이학박사학위논문

Identification of potential serum protein biomarkers for recurrence in gastric cancer patients using a quantitative multiple reaction monitoring approach

질량분석기 다중반응탐지법 기술을 통한

위암 재발 환자의 혈액 바이오마커 발굴 연구

2017년 07월

서울대학교 융합과학기술대학원

분자의학 및 바이오제약학과 단백체 생화학전공

조 병 규

Abstract

Identification of potential serum protein biomarkers for recurrence in gastric cancer patients using a quantitative multiple reaction monitoring approach

Byoung-Kyu Cho

Department of Molecular Medicine and Biopharmaceutical Sciences

Graduate School of Convergence Science and Technology

Seoul National University

Gastric cancer (GC) is one of the most common cancers representing the second leading cause of cancer-related mortality. Despite improvements in clinical therapies of GC, the recurrence rate of GC patients remains high (~55%) with advanced stage of the disease. Therefore, it is essential to understand of GC recurrence mechanisms that would help effective clinical

application for GC diagnosis and prognosis. Here, we aimed to identify potential serum biomarkers for recurrence in gastric cancers with an established quantitative multiple reaction monitoring (MRM) approach using GC patient serum samples. To build up a serum biomarker development platform, we first generated serum biomarker candidates through comprehensive proteomic approach. By employing both preliminary MRM and automated detection of inaccurate and imprecise transitions (AuDIT) analysis with stable isotope–labeled internal standard (SIS) peptides using pooled GC patient serum samples, we established a quantitative MRM analysis of 94 proteins as final MRM target proteins. To establish the multi-biomarker panel for identification of GC recurrence, we conducted the quantitative MRM analysis with 180 individual patients divided into the two groups, i.e. response group (n=133) and recurrence group (n=47), who received chemotherapy after D2 lymph node dissection in both groups, as a training set. By a stringent statistical analysis with quantitative MRM data of training set's individual samples, the 6-marker panel, consisting of alpha-1-antichymotrypsin (SERPINA3), apolipoprotein A-II (APOA2), apolipoprotein C-I (APOC1), clusterin (CLU), inter-alpha-trypsin inhibitor heavy chain H4 (ITIH4), and leucine-rich alpha-2-glycoprotein (LRG1), was constructed. These proteins showed the differentially expressed levels (p-value < 0.05) between the two groups with an area under the curve (AUC)

value of 0.810 and high prediction rates in both groups (95.5% and 61.7% in response and recurrence groups, respectively). To verify the 6-marker panel, we further applied MRM analysis with independent patient samples (n=64), i.e. response group (n=43) and recurrence group (n=21), as a test set. We demonstrated that 6 marker proteins showed the correlated expression patterns as in a training set with statistical significance (p-value < 0.05). We propose that these proteins can serve as diagnostic signatures to identify the recurrence in GC patients and our quantitative MRM assay based serum biomarker development platform could serve as a valuable tool in the clinical biomarker discovery-verification process.

Key words: Gastric cancer, Recurrence, Multiple Reaction Monitoring (MRM), Multi-biomarker panel

Student Number: 2013-30836

Abbreviations

ACN Acetonitrile, CH_3CN

AGC automatic gain control

CE Collisioin energy

CV Coefficient of variation

DTT Dithiothreitol

EOGC Early onset gastric cancer

FA Formic acid, HCO_2H

FDR False discovery rate

FWHM Full width at half maximum

GC Gastric cancer

HCD High energy collisional dissociation

HLC Hierarchical clustering

IAA Iodoacetamide

LC-MS/MS Liquid chromatography and tandem mass spectrometry

MRM Multiple reaction monitoring

MS Mass spectrometry

MS/MS Tandem mass spectrometry

NCE Normalized collision energy

QqQ Triple quadrupole

ROC Receiver operator characteristic

SIS Stable isotope–labeled internal standard

S/N Signal-to-noise ratio

Table of Contents

List of Tables

List of Figures

1. Introduction

Gastric cancer (GC) is the fourth most common cancer type representing the second leading cause of cancer-related mortality worldwide (Nagini 2012, Carcas 2014, Ferro, Peleteiro et al. 2014) and GC patients have a relatively poor prognosis, especially the patients of late pathologic stage with a less than 35% of 5-year survival rate (Yamazaki, Oshima et al. 1989). It has been reported that some risk factors of GC is associated with presence of Helicobacter pylori infection (Parsonnet, Friedman et al. 1991), age (>60 years), a history of stomach disorders or GC, accumulation of genetic alteration of multiple genes (Holian, Wahid et al. 2002), and the imbalance between cell proliferation and apoptosis (Zhou, Wong et al. 2001). Although there are some clinical diagnostic biomarkers in GC, such as the carcinoembryonic antigen (CEA), carbohydrate antigen 19–9 (CA 19–9), and carbohydrate antigen 72–4 (CA72–4), they are not specific and sensitive enough (Ebert and Röcken 2006) for monitoring the disease recurrence.

Despite improvements in modern curative surgery and/or postoperative chemo-radiotherapy including adjuvant treatments for GC, the majority of GC patients (~55%) suffer from local, regional or distant recurrence. In the early stages of GC patients are asymptomatic, and it is difficult to control the malignancy rate through early diagnosis and motivational therapy. Therefore,

it is important to find diagnostic biomarkers in GC that would help to understand of GC recurrence mechanisms and develop the effective clinical application.

Proteomics-based biomarker development platform using clinical patient's specimen sources has been rapidly evolved and can help to discover the accurate diagnostic and/or prognostic biomarkers (McDonald and Yates 2002, Schiess, Wollscheid et al. 2009, Frantzi, Bhat et al. 2014). Specifically, current advancements in multiple reaction monitoring (MRM) approach along with sensitive, selective, multiplexing and through-put analytical features could be able to implement for accurate quantification of biomarker candidates using triple quadrupole (QqQ) LC-MS instrumentation (Parker and Borchers 2014). Furthermore, a highly reproducible advance of MRM approach could result in reliable biomarkers that are clinically applicable for disease diagnosis and/or prognosis.

A general workflow of proteomics-based biomarker development platform is composed of two strategies including discovery and verification/validation stages. In discovery stage, a global proteomic profiling analysis is performed to determine the biomarker candidate proteins which show differentially expressed levels in control and case groups by a global proteomic profiling analysis. In validation stage, selected biomarker candidates are verified and/or

validated using the large cohort patient samples using the quantitative MRM analysis and statistical analysis (Figure 1).

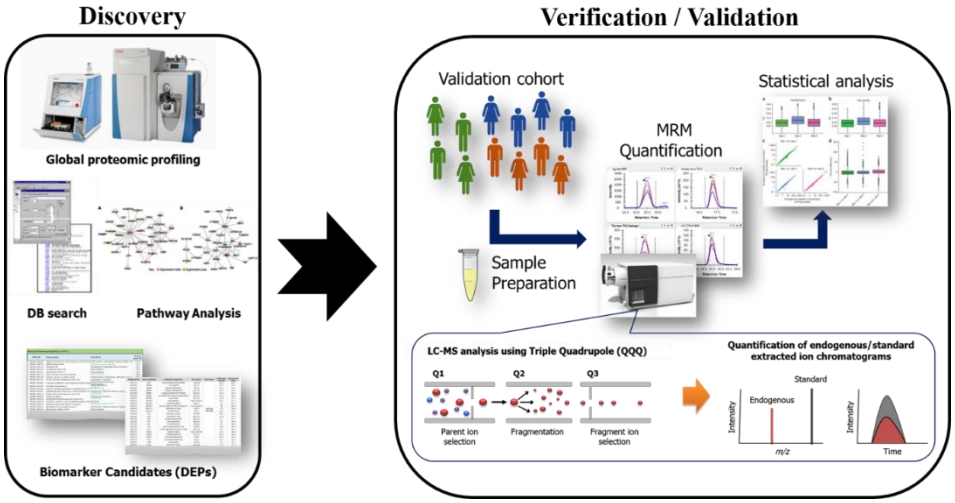

Figure 1. General workflow of proteomics-based biomarker development platform.

In the discovery, mass spectrometry analysis was performed to identify the differentially expressed proteins (DEPs) and then validated them by statistical and bioinformatics analysis tools to come up with the potential biomarker candidates. In verification and/or validation stage, selected biomarker candidates were verified using the large cohort patient samples by performing the MRM quantification and statistical analysis.

Most recently, several studies have attempted to investigate the potential diagnostic GC biomarkers using multiple proteomic strategies, such as matrix-assisted laser desorption/ionization (MALDI) analysis, isobaric tag for relative and absolute quantitation (iTRAQ) method, and MRM analysis. Using a combination of proteomic techniques, apolipoprotein C-I (APOC1), apolipoprotein C-III (APOC3), and prothrombin (F2) were identified as potential serum biomarkers for GC with both high sensitivity (>89.9%) and specificity (>71.0%) (Cohen, Yossef et al. 2011). In addition, four proteins including afamin (AFM), clusterin (CLU), vitamin D-binding protein (VTDB) and haptoglobin (HP) were detected as biomarkers for different stages of gastric cancer (Humphries, Penno et al. 2014). These studies demonstrate that proteomics-based biomarker development platform can identify the potential GC biomarkers by using a large cohort of clinical patient samples. Furthermore, several studies recently have attempted to investigate the GC recurrence mechanisms by discovering the early diagnostic biomarkers, however, the exact reasons and key factors are still elusive (D'Angelica, Gonen et al. 2004, Yong, Yang et al. 2012, Spolverato, Ejaz et al. 2014, Wu, Liu et al. 2014). Therefore, there is a need to find the prognostic biomarkers for monitoring the GC recurrence, which could ultimately help to extend GC patient's overall survival rate.

On the basis of the clinical significance of GC recurrence, we have established the serum biomarker development platform for identification of GC recurrence with MRM-based proteomic approach using GC patient serum samples. In this study, we especially applied to generate the multi-biomarker panel for overcome the limitations of single biomarkers, which often have inadequate predictive value and/or poor expression in particular patient samples. We discovered 94 biomarker candidates by integrative proteomic approach and verified them in the two groups; i.e. a training set (n=180) and a test set (n=64), using MRM analysis. To improve the discriminatory power, we constructed the 6-marker panel by combining several biomarker candidates which showed an AUC value of 0.810 and high prediction rates in two groups (95.5% and 61.7% in response and recurrence groups, respectively). Our quantitative MRM-based serum biomarker development platform enabled to verify the promising multi-biomarker panel of GC recurrence which may contribute to identify the GC response and recurrence patients by prediction of their recurrence possibility.

2. Material and Methods

2.1. Reagents

HPLC grade acetonitrile (ACN) and water were purchased from J.T Baker (Phillipsburg, NJ, USA). Formic acid (FA), urea, dithiothreitol (DTT), iodoacetamide (IAA), and ammonium bicarbonate were purchased from Sigma-Aldrich (St. Louis, MO, USA). Sep-pak C18 cartridges were from Waters (Milford, MA, USA). Sequencing grade modified porcine trypsin were obtained from Promega (Madison, WI, USA). SIS peptides containing a single amino acid labeled with ^{13}C and ^{15}N were synthesized at crude levels from JPT (Berlin, Germany).

2.2. Study population

Between November 2004 and April 2008, 248 GC patients were recruited to the study at Samsung Medical Center (Seoul, Korea). The following criteria were adopted for patient groups in this study; 1) histologically confirmed adenocarcinoma of the stomach, 2) D2 lymph node dissection, 3) Eastern Cooperative Oncology Group performance status of 0 or 1, 4) adequate function of major organs including cardiac, hepatic, and renal, and 5) adequate bone marrow function (hemoglobin > 10g/㎗; absolute neutrophil count [ANC]

≥ 1,500/㎕; platelet count ≥ 100,000/㎕). Patients with coexisting malignancies or who were unable to tolerate chemotherapy due to other systemic illnesses were excluded from the study. Eligible 248 GC patients were randomly assigned to receive the adjuvant chemotherapy with capecitabine plus cisplatin (XP) and XP plus radiotherapy with capecitabine (XPRT). The XP treatment was repeated every 3 weeks with six cycles of capecitabine 1,000 mg/m^2 twice a day on days 1 to 14 and cisplatin 60 mg/m^2 on day 1. GC patients assigned to XPRT treatment, radiotherapy 45 Gy concurrently with 825 mg/m^2 twice a day was given after the completion of two cycles of XP, followed by two additional cycles of XP. Radiotherapy was fractionated to 1.8 Gy daily, 5 days a week, over 5 weeks. All 248 patients were adhered to the same schedule of the median follow-up duration of 7 years after completion of study treatment. Disease-free survival (DFS) was measured from the date of surgery until death, recurrence, second primary tumor, whichever occurred first. All patient serum samples were collected 4-week after D2 lymph node dissection (Park, Sohn et al. 2015).

2.3. Serum sample preparation

All 248 patient's blood samples were centrifuged immediately at 2,000 g for

10 minutes to fractionate the serum. The resulting supernatant was aliquoted (50 – 100 ㎕) and stored at -80℃ until analysis. For global proteomic analysis, serum samples were immunodepleted of fourteen high-abundant proteins (albumin, antitrypsin, IgA, IgG, IgM, transferrin, haptoglobin, fibrinogen, alpha2-macroglobulin, alpha1-acid glycoprotein, apolipoprotein AI, apolipoprotein AII, complement C3, and transthyretin) using a MARS Hu-14 column (Agilent Technologies, Santa Clara, CA) equipped with an Ultimate 3000 HPLC system (Thermo Fisher Scientific, San Jose, CA). Crude serum samples were diluted with buffer A and filtered through a 0.22 ㎛ spin filter at 16,000 g at room temperature for 2 minutes to remove particulates. Diluted serum samples were injected at 0.125 ml/min and flow-through fraction containing the low-abundant proteins was collected. Depleted serum samples were concentrated by a 3 kDa molecular weight cutoff (MWCO) filter. Depleted serum protein concentration was measured using the Micro BCA protein assay kit (Thermo Scientific, Rockford, IL). Equal amounts of proteins from each group were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) with 4 ~ 12% Bis-Tris Gel (Novex NUPAGE electrophoresis system, MES running buffer) (Invitrogen, Carlsbad, CA, USA) and stained with Coomassie Brilliant Blue (Sigma-Aldrich). Each gel lane was cut into nine pieces and subjected to in-gel tryptic digestion following the

general protocol. Briefly, excised protein bands were destained, reduced with 20 mM DTT at 60℃ for one hour and then alkylated with 55 mM iodoacetamide at room temperature for 45 minutes in the dark. After dehydration with ACN, the proteins were digested with 13 ng/㎕ sequencing grade modified trypsin (Promega, Madison, WI, USA) in 50 mM ammonium bicarbonate overnight at 37° C. Peptides were extracted from the gel slices with 50% (v/v) ACN in 5% (v/v) formic acid (FA). The eluates were dried under a Centrivap concentrator (Labconco, Kansas City, MO) and stored at -20°C until used.

For MRM-MS analysis, serum samples were immunodepleted of six high-abundant proteins (albumin, IgA, IgG, transferrin, haptoglobin, and antitrypsin) using a MARS Hu-6 column equipped with HPLC system as previoulsy described. Protein was reduced with 6M urea and 10mM DTT, and alkylated with 30 mM iodoacetamide. The sample was then diluted to 1M urea with 50mM ammonium bicarbonate, and sequencing grade modified trypsin was added 1:50 (trypsin:protein) ratio and incubated overnight at 37°C. Peptides were eluted with 50% (v/v) ACN in 0.1% (v/v) FA and then desalted using Sep-pak C18 cartridges. Desalted peptides were lyophilized under a Centrivap concentrator (Labconco, Kansas City, MO) and stored at -80℃

until use.

2.4. Mass spectrometry analysis

Extracted peptide samples were suspended in 0.1% FA and separated on a PepMap^{TM} RSLC C18 column (Thermo Fisher Scientific, San Jose, CA) with a linear gradient of 2~38% solvent B (0.1% FA in ACN) in for 160 minutes at a flow rate of 300 nl/min. Samples were analyzed in duplicate on a Q-Exactive (Thermo Fisher Scientific, San Jose, CA) hybrid quadrupole-Orbitrap mass spectrometry, interfaced with Ultimate 3000 HPLC system (Thermo Fisher Scientific, San Jose, CA). The spray voltage was set to 2.0 kV and the temperature of the heated capillary was set to 250°C. The Q-Exactive was operated in a data-dependent mode with one survey MS scan followed by ten MS/MS scans and a dynamic exclusion time of 30 seconds. The full scans were acquired in the mass analyzer at 300 ~ 1400 *m/z* with the resolution of 70,000 and the MS/MS scans were obtained with the resolution of 17,500 by using a normalized collision energy (NCE) of 27% for high energy collisional dissociation (HCD) fragmentation. The automatic gain control (AGC) target was set to $1e^5$, maximum injection time was 120 ms and the isolation window was set to 2 *m/z*.

2.5. Database search

Collected MS/MS data were converted into mzXML files through the Trans Proteomic Pipeline (version 4.5) software and searched against the decoy uniprot Human database (June 2014, 313072 entries) for the estimation of the false discovery rate (FDR) with the SEQUEST (Thermo Fisher Scientific, version 27) program in the Proteome Discoverer™ (Thermo Fisher Scientific, version 1.4) search platform. Full tryptic specificity and up to two missed cleavage sites were allowed; mass tolerances for precursor ions and fragment ions were set to 10 ppm and 0.5 Da, respectively; fixed modification for carbamidomethyl-cysteine and variable modifications for methionine oxidation were used. The Scaffold software package (version 4.6.2, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identification. All proteins with a ProteinProphet probability ≥ 99% and a PeptideProphet probability ≥ 95% were identified (Nesvizhskii, Keller et al. 2003).

2.6. Relative protein quantification

Relative protein quantification was accomplished using the label-free technique, spectral counting. The MS/MS data were normalized to compare abundances of proteins between two groups using Scaffold software. The normalized spectral counts from duplicate analyses of a control group and an

experimental group were compared using the R program (version 2.15) with power law global error model (PLGEM) software (http://www.bioconductor.org), a statistical analysis software package in order to identify DEPs with statistically significant protein changes (signal-to-noise ratio (S/N) and *p*-value) (Pavelka, Fournier et al. 2008).

2.7. Selection of MRM peptide transitions

MRM peptide transitions were determined using two scan modes; Unbiased Q3-ion monitoring (Cho, Koo et al. 2014)**:** The unbiased Q3-ion monitoring assay was performed using MRM mode with the following steps; 1) most intense five fragment y and/or b ions were selected from both Q-exactive MS/MS spectra and PeptideAtlas database of target peptides, 2) dwell time of each transition was set to 5 ms and Q1/Q3 transitions were monitored with a unit resolution mass window (0.7 FWHM, full width at half maximum), and 3) chromatographic elution profiles of each transition ion and estimation of chromatographic peak area were carried out using Mass Hunter Quantitative Analysis (MHQA) software (version B.6.0, Agilent Technologies). Agilent's MassHunter Optimizer software: An Agilent's MassHunter Optimizer software was utilized with synthetic peptides with the following steps; 1) doubly and triply charged precursor ions were selected from each target peptide, 2) theoretical fragment y ions were selected from each target peptide,

3) selected Q1/Q3 transitions were monitored with ten different collision energies (5 ~ 50 V) using MRM mode, and 4) estimation of chromatographic peak area were carried out using MHQA software and most intense five transitions were selected.

2.8. AuDIT analysis for determination of accurate and precise MRM transitions

A total of 228 SIS peptides corresponding initial MRM target peptides were synthesized at 12 nmol. Stock solutions were prepared at 100 pmol/$\mu\ell$ with 30% ACN/0.1% FA and further diluted to 200 fmol/$\mu\ell$ with 0.1% formic acid. MRM analysis was performed in triplicate with a total of 1824 MRM transitions using an equimolar mixture (50fmol) of the 228 SIS peptides spiked into 5 μg of digested pooled serum samples. For the 3-MRM data set, data integration and AuDIT analysis were performed using Skyline and QuaSAR program. Briefly, AuDIT uses the p-value of t-test and the coefficient of variation (CV) of the ratio of analyte's peak area ratio (PAR) to SIS's to detect problem transitions. Final MRM transitions were determined by two thresholds (p-value < 10^{-5} for t-test and CV $\leq$ 20% for PAR, respectively) (Abbatiello, Mani et al. 2010, Burgess, Keshishian et al. 2014).

2.9. Functional enrichment of MRM target proteins by STRING database

To evaluate functional enrichment and interactions between final MRM target proteins, we used STRING database (version 10.5, http://string-db.org). It allows to assign the biological process, molecular function, cellular component, and KEGG pathways by FDR scores based on experimental repositories, public text collections, and computational prediction methods (Szklarczyk, Franceschini et al. 2014).

2.10. Study design of MRM analysis order with sample randomization

Randomization of analyzed sample order can prevent the unbiased systematical effects for improving molecular biomarker discovery (Qin, Zhou et al. 2014). Randomization for injection order of MRM analysis with 180 individual patient's samples was performed using True Random Number Service (www.random.org).

2.11. Multiple reaction monitoring analysis

Peptides were suspended in 0.1% FA and separated on a ZORBAX Eclipse Plus C18 column (2.1x150mm id., 1.8-㎛ particle size, Agilent) with a linear

gradient of 0~40% solvent B (0.1% FA in ACN) in for 50 minutes at a flow rate of 250 μℓ/min. Samples were analyzed on an Agilent 6490 triple quadrupole mass spectrometer (Agilent Technologies) interfaced with an Agilent 1290 Infinity HPLC system.

The ion spray capillary voltage and nozzle voltage set to 3500 and 500 V, respectively. The drying gas temperature was set to 250°C at 12 L/min, and the sheath gas temperature was 350°C at 11 L/min. The nebulizer was set to 50 psi. The fragmentor voltage and the cell accelerator voltage was set to 380 V and 5 V, respectively. Both precursor and product ions were monitored with a unit resolution mass window (0.7 FWHM) in Q1 and Q3. In the dynamic MRM measurement, transitions were acquired during for retention time ±4 min of corresponding peptide. Chromatographic elution profiles of each transition ion and estimation of chromatographic peak area were achieved using both MHQA and Skyline software.

2.12. Statistical analysis

To identify the significant DEPs in two analyzed sample groups, we performed a statistical analysis using MSstats (version 3.8.2) (Surinova, Hüttenhain et al. 2013, Choi, Chang et al. 2014). In the first step, data preprocessing was performed by transforming all transition intensities into $\log_2$ values. Then, a

constant normalization was conducted on the basis of reference SIS peptide's transitions for all target peptides, which equalized the median peak ratio intensities of reference transitions from all target peptides across all MRM runs and adjusted the bias to SIS peptide's signals. Protein-level quantification and testing for differential abundance in both response and recurrent groups were performed with the linear mixed model (LMM) implemented in MSstats. Each protein is tested for abundant differences between two groups. The *p*-values were adjusted to control the FDR at a cutoff of 0.05. All proteins with a *p*-value below 0.05 were considered significant.

Hierarchical clustering (HLC) analysis was conducted with log_2-transformed peak area's ratio of significant DEPs to confirm their distribution patterns using Multi Experiment Viewer (Version 4.9, http://www.tm4.org/mev/).

The stepwise logistic regression analysis was performed to generate the multi-marker panel which contributed to the discriminatory power between two analyzed groups (Klecka 1980). The receiver operator characteristic (ROC) curves and interactive plots were generated based on DeLong's method (DeLong, DeLong et al. 1988). All analysis was conducted using MedCalc software (Mariakerke, Belgium, version 12.2.1.).

2.13. Evaluation of quantitative linear curves

Six SIS peptides (EQLTPLIK of APOA2, EFGNTLEDK of APOC1,

ASSIIDELFQDR of CLU, ITFELVYEELLK of ITIH4, DLLLPQPDLR of LRG1, and ITLLSALVETR of SERPINA3, respectively) were spiked into 5 μg of digested pooled serum samples and serially diluted at 10 concentrations (1, 2.5, 5, 10, 25, 50, 100, 250, 500, and 1000 fmol/μℓ). All MRM assay were analyzed in triplicate and the normalized peak area was used to generate linear response curves.

3. Results

3.1. Clinicopathologic characteristics of study population

In this study, we selected a total of 248 GC patients who received D2 lymph node dissection at the Samsung Medical Center (Seoul, Korea). All patients were randomly assigned, 107 patients (43.1%) to the XP treatment and 141 patients (56.9%) to the XPRT treatment as described in Material and Method 2.2 (study population). After median follow-up of 7 years, 70 recurrence events (28.2%) occurred and the estimated 5-year DFS rates were 69.0% (172 patients alive). In total patient population, 58 (23.4%) patients, 88 (35.5%) patients, 70 (28.2%) patients and 32 (12.9%) patients were diagnosed as stage I, stage II, stage III and stage VI, according to the Joint Committee on Cancer Staging System (six edition, 2002), respectively. Clinicopathologic characteristics of study population are summarized in Table 1.

To generate the multi-marker panel, we divided a total of 248 GC patients into 2 sub-cohorts which were training and test sets. When we selected the number of patients in each set, we performed sample size calculation using MedCalc software, taking into account the required significance level and power of the test based on two required inputs; 1) the hypothesized AUC value over 0.8 and 2) ratio of sample sizes in negative and positive groups (recurrence group : response group = 1 : 2.5). As a result, we selected 184 patients as a training

set to construct of multi-marker panel and other 64 patients as a test set to evaluate the reliable performance of the multi-biomarker panel.

Table 1. Baseline patient demographics and clinical characteristics.

	Response group (n=135)	Recurrence group (n=49)
Age		
Median (range)	50 (26-75)	54 (34-80)
Sex		
Male / Female (Male %)	120/58 (67.4%)	53/17 (75.7%)
Assigned chemotherapy, n (%)		
XP treatment	78 (43.8%)	29 (41.4%)
XPRT treatment	100 (56.2%)	41 (58.6%)
Primary tumor location, n (%)		
Cardia	8 (4.5%)	3 (4.3%)
Fundus	1 (0.6%)	0 (0.0%)
Body	100 (56.1%)	30 (42.9%)
Antrum	68 (38.2%)	36 (51.4%)
Angle	1 (0.6%)	1 (1.4%)
***Pathological stage, n (%)**		
IB	54 (30.3%)	4 (5.7%)
II	74 (41.6%)	14 (20.0%)
III	42 (23.6%)	28 (40.0%)
IV	8 (4.5%)	24 (34.3%)

* According to the Joint Committee on Cancer Staging System, six edition (2002).

3.2. Overall scheme of serum biomarker development for gastric cancer recurrence using proteomic approach

In this study we aimed to find multi-biomarker panel for GC recurrence patients who received chemotherapy, i.e. XP treatment or XPRT treatment, after D2 lymph node dissection. As shown in Figure 2, we first selected a set of 330 proteins as initial biomarker candidates via integrative proteomic approach using three independent strategies (global profiling, public database, and in-house dataset from EOGC). We then performed the preliminary MRM and AuDIT analysis to determine final MRM targets which showed a reliable detectability with an accurately and precisely quantitative manner in our QqQ LC-MS system. As a result, 135 peptides derived from 94 proteins were determined as final biomarker candidates for identification of recurrence in gastric cancers using MRM analysis. To generate the multi-biomarker panel for identification of GC recurrence, we divided all individual patient samples in two groups; 1) GC patients who subsequently developed the recurrence events as a recurrence group and 2) their controls who showed survival outcome as a response group compared to the recurrence group.

We first applied multiplexed-MRM analysis with a training set (n=180) and measured the differentially expressed levels of 94 biomarker candidates between the response (n=133) and recurrence (n=47) groups. We identified 65

DEPs which showed the quantitative differences in the two groups with a statistically significant value (p-value < 0.05) by MSstats analysis. We then constructed multi-biomarker panel by combination of these 65 DEPs to obtain the best model which could discriminate the two groups. Finally, we further performed multiplexed-MRM analysis with a test set (n=64) to demonstrate if the constructed multi-biomarker panel had a reproducible performance to distinguish the two groups as exhibited in training set's results.

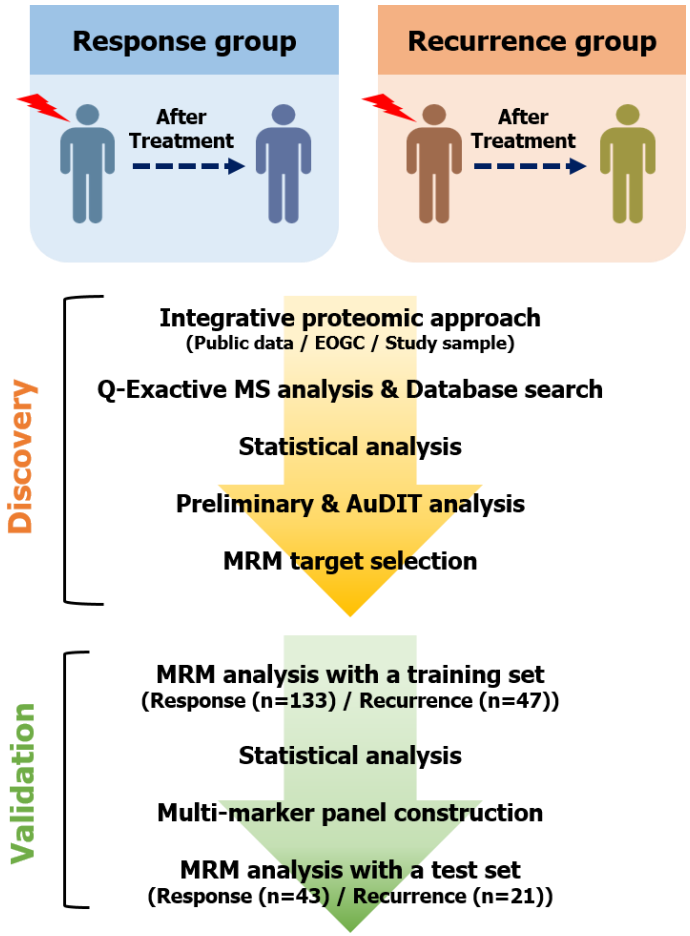

Figure 2. Overall scheme of serum biomarker development for identification of recurrence in gastric cancers.

In the discovery stage, initial biomarker candidates for identification of recurrence in gastric cancers were selected via integrative proteomic approaches including public data mining, EOGC data mining, and proteomic profiling of study GC patient serum samples. Next, the preliminary and AuDIT MRM analysis were performed to determine final MRM target proteins

showing the reliable quantitative results in MRM analysis.

In the validation stage, multiplexed-MRM analysis was performed with the training set (n=180) and constructed a multi-biomarker panel by statistical analysis. Finally, independent test set (n=64) was analyzed by multiplexed-MRM experiment to verify the constructed multi-biomarker panel.

3.3. Determination of potential serum biomarker candidates for identification of recurrence in gastric cancers

To establish the serum biomarker discovery platform for identification of GC recurrence using MRM analysis, we first determined reliable biomarker candidates which may have the potential prognosis for GC recurrence by a comprehensive proteomic approach. This strategy is composed of three independent approaches as follows; 1) a knowledge-based approach with global data mining of previously reported protein biomarkers involved in GC, 2) a proteomic profiling analysis using early onset gastric cancer (EOGC) patient samples who may suffer GC recurrence with an increased risk due to their genetic abnormality, and 3) a global proteomic profiling analysis using pooled study serum samples (n=2) who showed extremely DFS values from each group to identify the differentially expressed serum proteins between the two groups.

1) It has been well documented that a global data mining with previously reported biomarker proteins and/or proteomic data is a prerequisite approach for evaluation of proteomics-based biomarker discovery platform (Killcoyne, W Deutsch et al. 2012, Younesi, Toldo et al. 2012, Griss, Perez-Riverol et al. 2015). We screened the known potential biomarkers for GC with numbers of literatures which have been published after 2009 (Qiu, Yu et al. 2009, Lastraioli, Raffaella Romoli et al. 2012, Li, Zhang et al. 2012, Lin, Huang et

al. 2012, Leal, Assumpção et al. 2014, Tsai, Wang et al. 2014, Wu, Cheng et al. 2014, Elimova, Wadhwa et al. 2015, Jin, Jiang et al. 2015, Mi, Ji et al. 2016). In addition, we obtained biomarker candidates from the public database (A Database of Human Gastric Cancer, DBGC) which integrates various human GC related data resources (Wang, Zhang et al. 2015) (http://bminfor.tongji.edu.cn/dbgc/index.do). Through these approaches, we found that 306 proteins were previously reported as GC-associated proteins. Since our aim was to investigate serum biomarker proteins for identification of GC recurrence, we then sorted out 306 proteins using both Plasma Proteome database (http://www.plasmaproteomedatabase.org/) and Secreted Protein Database (http://spd.cbi.pku.edu.cn/). As a result, 146 serum proteins were selected potential biomarker candidates for GC recurrence.

2) Early onset gastric cancer (EOGC) is defined as gastric cancer presenting at the age of 45 or younger (Milne, Sitarz et al. 2007, Milne and Offerhaus 2010). Although rare proportion (approximately 10%) presenting with GC states earlier than 45 years, it is well known that EOGC patients usually show poor outcome with GC recurrence (Yoo, Noh et al. 2000, Otsuji, Kuriu et al. 2004, Kim, Lee et al. 2014). Since the genetic alterations of EOGC patients present different patterns compared to older patients, it is importantly thought that unraveling genetic mechanisms of EOGC can provide vital information of molecular pathway associated with GC (Maehara, Emi et al. 1996, Rugge,

Busatto et al. 1999, Haruma, Komoto et al. 2000). In this study, we selected biomarker candidates for GC recurrence from previously global proteomic profiling data using tumor and adjacent normal tissues collected from three different EOGC patients (Park, Park et al. 2015). The proteins with p-value < 0.05 at least in one of the three tissue pairs were identified as DEPs and a total of 118 DEPs were determined as biomarker candidates for GC recurrence.

3) Although it is important to take advantage of previously reported biomarkers for design of biomarker development study using proteomic approach, the most desirable way is to utilize actual experimental LC-MS/MS profiling datasets from the given biological sample matrices because the detectability of biomarker candidates is a critical factor for accurate quantification by MRM analysis. In this study, we employed global proteomic profiling by LC-MS/MS analysis using study serum samples obtained from GC response and recurrence patient groups to identify the DEPs between two analyzed groups (Figure 3). Initially, we selected two patient samples from each group with representing extremely different disease free survival (DFS) values to observe the definite quantitative abundances of serum proteins in the two groups. We then carried out the serum-pooling strategy (n=2) of each group, followed by the depletion of highly abundant proteins, SDS-PAGE protein fractionation, and mass spectrometry analysis. Briefly, an equal

volume (30 μℓ) of two patient serum samples from each group was pooled separately and immunodepleted of fourteen interfering high abundant proteins in order to identify serum proteome with an effectively expanding of the dynamic range. To improve proteome coverage and identify low abundance proteins, the depleted serum proteins (30 μg) from each group were subsequently separated on an SDS-PAGE gel and divided into 9 fractions (Figure 4) followed by duplicate LC-MS/MS analysis. We identified a total of 846 proteins (95.0% peptide probability, 99.0% protein probability, minimum one peptide, peptide FDR 0.3%) with the Proteome Discoverer software (Version 1.4) using a SEQUEST search engine (Figure 5). In addition, we compared the biological processes of identified serum proteins between the two groups using Scaffold (version 4.6.2) (Figure 6).

To estimate the significant DEPs between the two groups, we normalized spectral counts in both groups using Scaffold and further statistically validated by PLGEM analysis to obtain signal-to-noise ratio and *p*-values for each protein quantification. As a result, we identified 82 DEPs that satisfied the cutoff for the signal-to-noise ratio (STN ≥2) (Table 2). Among the 82 proteins, 54 proteins were significantly up-regulated and 28 proteins were down-regulated in GC recurrence group, as in their expression levels in response

group.

Based on our three independent approaches, a total of 330 proteins were selected as initial serum biomarker candidates for identification of recurrence in GC patients.

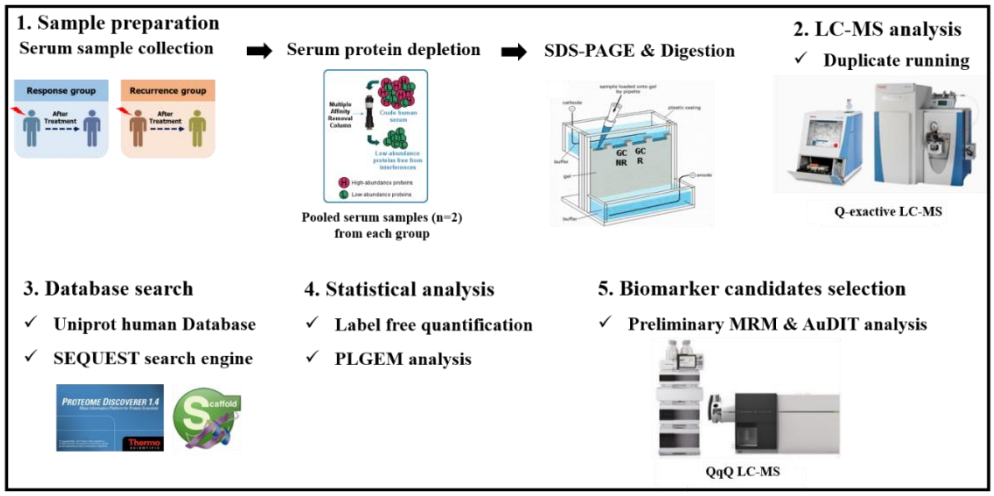

Figure 3. Overall workflow of a global proteomic profiling analysis using study samples.

First, two patient samples from each group were selected with DFS values and then carried out the serum-pooling strategy, the depletion of fourteen highly abundant proteins, and SDS-PAGE protein fractionation. Next, duplicate LC-MS/MS analysis was performed with Q-exactive LC-MS system and database search was conducted with the Proteome Discoverer software using a SEQUEST search engine. Expression levels of identified proteins were normalized with their spectral counts using Scaffold and statistically validated by PLGEM analysis to obtain signal-to-noise ratio and *p*-values to estimate the significant DEPs in the two groups. Finally, initial biomarker candidates were selected by both preliminary and AuDIT analysis.

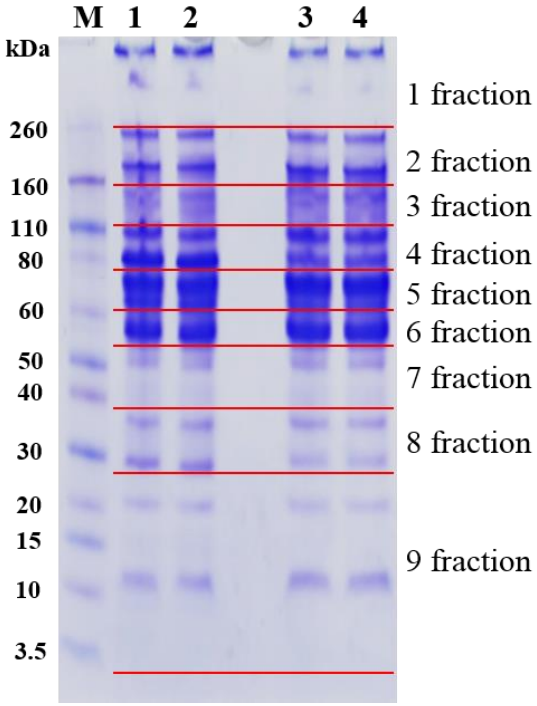

Figure 4. SDS-PAGE fractionation of immunodepleted serum proteins.

Pooled immunodepleted serum proteins from adjuvant therapy response patients (n=2) and recurrence patients (n=2) were separated on a 4 ~ 12% Bis-Tris gel. After staining with Coomassie brilliant Blue, gel lanes were cut into nine pieces as indicated and subjected to in-gel tryptic digestion. Protein markers (lane M), serum samples of recurrence group (each 30 μg in lane 1 and 2, respectively), and serum samples of response group (each 30 μg in lane 3 and 4, respectively).

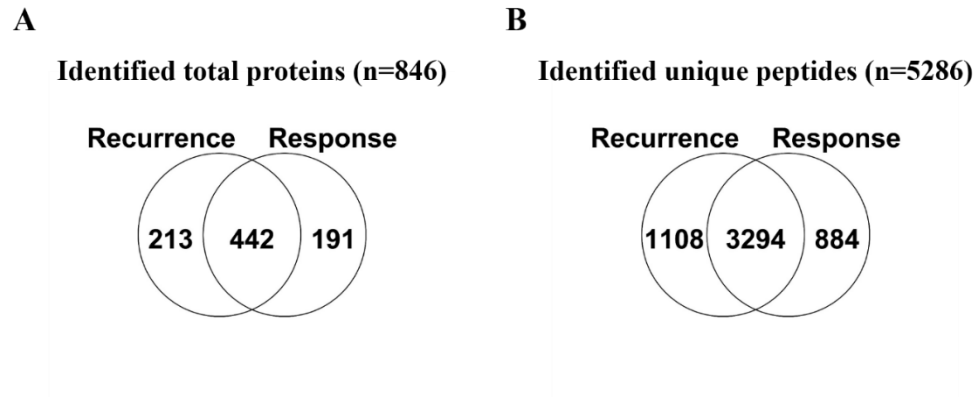

Figure 5. Venn diagrams of the numbers of identified proteins and peptides by LC-MS/MS analysis.

(A) A total of 846 serum proteins and (B) their corresponding 5286 unique peptides were identified by duplicate LC-MS/MS analysis from both recurrence and response groups.

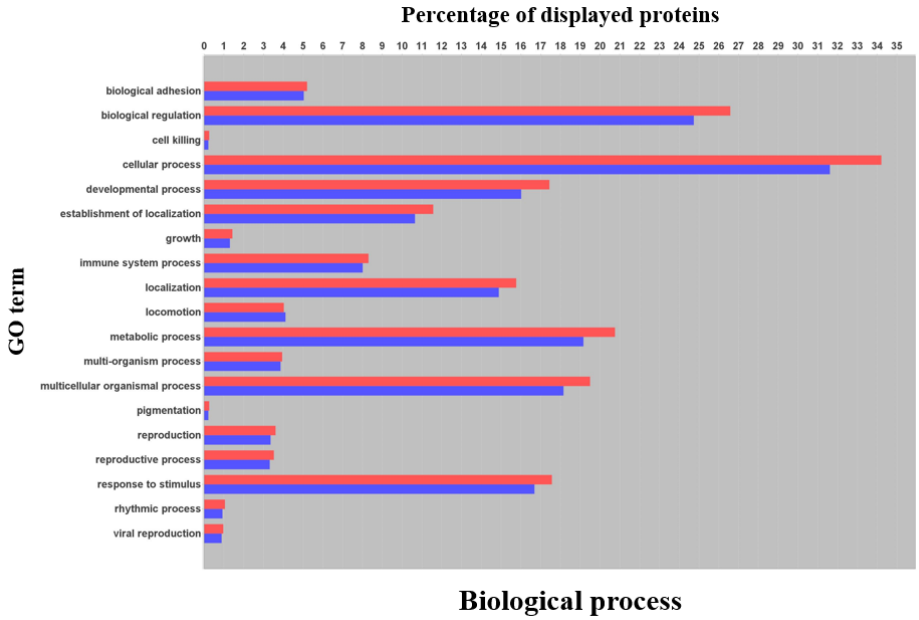

Figure 6. Functional enrichment of identified serum proteins.

A total of 846 serum proteins obtained from the two groups were classified into biological process subcategories using Scaffold software. Percentage of displayed proteins from both recurrence and response groups are shown in red and blue, respectively.

Table 2. Differentially expressed serum proteins in GC recurrence group.

Uniprot ID	Gene name	Protein	*STN	*p*-value
P01009	SERPINA1	Alpha-1-antitrypsin	60.7	0
P01008	SERPINC1	Antithrombin-III	28.5	0
P02774	GC	Vitamin D-binding protein	25.8	0
P02790	HPX	Hemopexin	25.7	0
P0C0L5	C4B	Complement C4-B	23.3	0
P00450	CP	Ceruloplasmin	21.8	0
P01024	C3	Complement C3	17.5	0
P01011	SERPINA3	Alpha-1-antichymotrypsin	17.5	0
P01042	KNG1	Kininogen-1	11.7	0
P02746	C1QB	Complement C1q subcomponent subunit B	11.0	0
P36955	SERPINF1	Pigment epithelium-derived factor	10.9	0
P02749	APOH	Beta-2-glycoprotein 1	10.4	0
P02747	C1QC	Complement C1q subcomponent subunit C	10.3	0
Q06033	ITIH3	Inter-alpha-trypsin inhibitor heavy chain H3	10.3	0
P00738	HP	Haptoglobin	8.8	0
P01031	C5	Complement C5	8.5	0
P02743	APCS	Serum amyloid P-component	7.4	0
P04114	APOB	Apolipoprotein B-100	7.2	0
P10643	C7	Complement component C7	7.0	0
P02748	C9	Complement component C9	6.8	0
P13671	C6	Complement component C6	6.6	0
P02741	CRP	C-reactive protein	6.0	0
P15169	CPN1	Carboxypeptidase N catalytic chain	5.8	0
Q14624	ITIH4	Inter-alpha-trypsin inhibitor heavy chain H4	5.7	0
P08603	CFH	Complement factor H	5.5	0
P02765	AHSG	Alpha-2-HS-glycoprotein	5.4	0
Q96PD5	PGLYRP2	N-acetylmuramoyl-L-alanine amidase	4.8	0.001
E7ETH0	CFI	Complement factor I	4.7	0.001
P06396	GSN	Gelsolin	4.4	0.001
P05543	SERPINA7	Thyroxine-binding globulin	4.3	0.001
P02745	C1QA	Complement C1q subcomponent subunit A	4.2	0.001

P04217	A1BG	Alpha-1B-glycoprotein	4.2	0.001
P02750	LRG1	Leucine-rich alpha-2-glycoprotein	4.0	0.001
P07360	C8G	Complement component C8 gamma chain	4.0	0.001
P07358	C8B	Complement component C8 beta chain	4.0	0.001
A0A087X2C0	IGHM	Ig mu chain C region	3.9	0.001
P05546	SERPIND1	Heparin cofactor 2	3.3	0.002
P00747	PLG	Plasminogen	3.2	0.003
P22352	GPX3	Glutathione peroxidase 3	3.0	0.003
Q9Y6R7	FCGBP	IgGFc-binding protein	3.0	0.003
P29622	SERPINA4	Kallistatin	2.9	0.003
P07357	C8A	Complement component C8 alpha chain	2.8	0.004
O00187	MASP2	Mannan-binding lectin serine protease 2	2.7	0.004
Q8NBP7	PCSK9	Proprotein convertase subtilisin/kexin type 9	2.7	0.004
P02766	TTR	Transthyretin	2.6	0.004
P36980	CFHR2	Complement factor H-related protein 2	2.5	0.005
P19827	ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1	2.5	0.006
P22792	CPN2	Carboxypeptidase N subunit 2	2.4	0.007
Q96IY4	CPB2	Carboxypeptidase B2	2.4	0.007
P01877	IGHA2	Ig alpha-2 chain C region	2.4	0.007
P68871	HBB	Hemoglobin subunit beta	2.4	0.007
P29320	EPHA3	Ephrin type-A receptor 3	2.3	0.007
Q6EMK4	VASN	Vasorin	2.2	0.008
P0C0L4	C4A	Complement C4-A	2.0	0.008
P35222	ACTB	Actin, cytoplasmic 1	-2.1	0.008
P80108	GPLD1	Phosphatidylinositol-glycan-specific phospholipase D	-2.2	0.007
P14543	NID1	Nidogen-1	-2.2	0.007
P43652	AFM	Afamin	-2.4	0.006
P02652	APOA2	Apolipoprotein A-II	-2.4	0.006
P02775	PPBP	Platelet basic protein	-2.5	0.005
P70232	CHL1	Neural cell adhesion molecule L1-like protein	-2.6	0.004
P02649	APOE	Apolipoprotein E	-2.6	0.004
P49747	COMP	Cartilage oligomeric matrix protein	-3.0	0.003
P04275	VWF	von Willebrand factor	-3.1	0.003
P03952	KLKB1	Plasma kallikrein	-3.1	0.003

P01860	IGHG3	Ig gamma-3 chain C region	-3.3	0.003
O75473	LGR5	Leucine-rich repeat-containing G-protein coupled receptor 5	-3.5	0.002
O75882	ATRN	Attractin	-3.9	0.001
P00739	HPR	Haptoglobin-related protein	-4.0	0.001
A2MYD4	V2-7	V2-7 protein	-4.1	0.001
P02753	RBP4	Retinol binding protein 4,	-6.3	0
Q8N5F4	IGL@	IGL@ protein	-6.3	0
P04004	VTN	Vitronectin	-6.7	0
P04196	HRG	Histidine-rich glycoprotein	-6.7	0
P19823	ITIH2	Inter-alpha-trypsin inhibitor heavy chain H2	-8.9	0
P07996	THBS1	Thrombospondin-1	-9.7	0
P01023	A2M	Alpha-2-macroglobulin	-10.8	0
P10909	CLU	Clusterin	-12.1	0
P02787	TF	Transferrin	-14.5	0
P25311	AZGP1	Zinc-alpha-2-glycoprotein	-15.0	0
P02751	FN1	Fibronectin 1	-16.9	0
P02647	APOA1	Apolipoprotein A-I	-30.9	0

*STN : Signal-to-noise ratio generated by PLGEM analysis.

3.4. Determination of MRM targets by both preliminary MRM and AuDIT analysis

A key step for quantitative MRM analysis is the determination of confident and detectable MRM transitions within a QqQ LC-MS system (Gerber, Rush et al. 2003, Kirkpatrick, Hathaway et al. 2006). In this study, a set of MRM candidate peptides (≥5) for each target protein were selected from PeptideAtlas DB (http://www.peptideatlas.org/) following the general selection criteria, such as uniqueness to the target proteins, the length of peptides (10 ~ 25 amino acids) and the charge states that can be detected within the *m/z* scan window. Next, we selected MRM transition candidates (≥5) as their intensity order of each target peptide using both PeptideAtlas DB and our global proteomic MS/MS data. With these MRM target candidates, we performed preliminary MRM analysis to confirm their reliable detectability using the unbiased Q3 ion monitoring method in our QqQ LC-MS system (Cho, Koo et al. 2014) (Figure 7). As a result, 228 peptides derived from 141 proteins were selected our initial MRM targets following the selection criteria; at least 5 transitions were detected per target peptide with co-eluted chromatographic patterns as showing the S/N ratio above 3 (Figure 8).

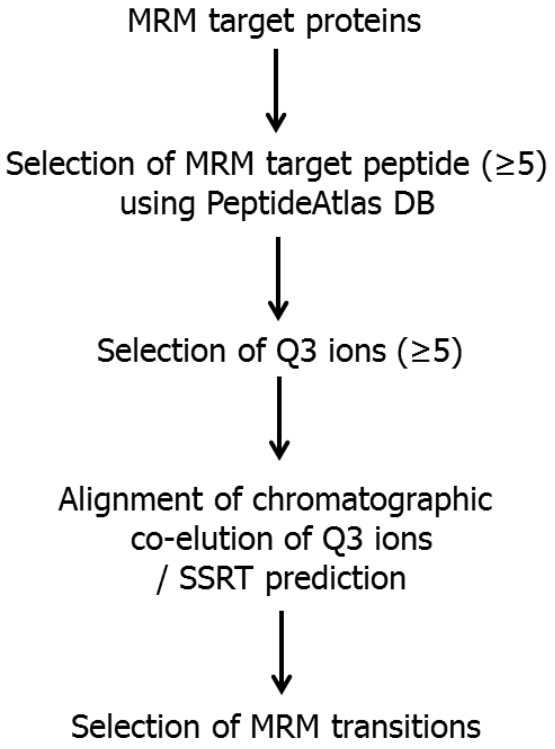

Figure 7. Workflow for determination of MRM transitions.

MRM target peptides (≥5) for each target protein were selected by PeptideAtlas DB. Next, the MRM transitions of target peptides were determined by the unbiased Q3 ion monitoring using QqQ LC-MS (Cho, Koo et al. 2014). All transitions were confirmed by co-elution patterns of Q3 ions and correlation with Sequence Specific Retention Time (SSRT).

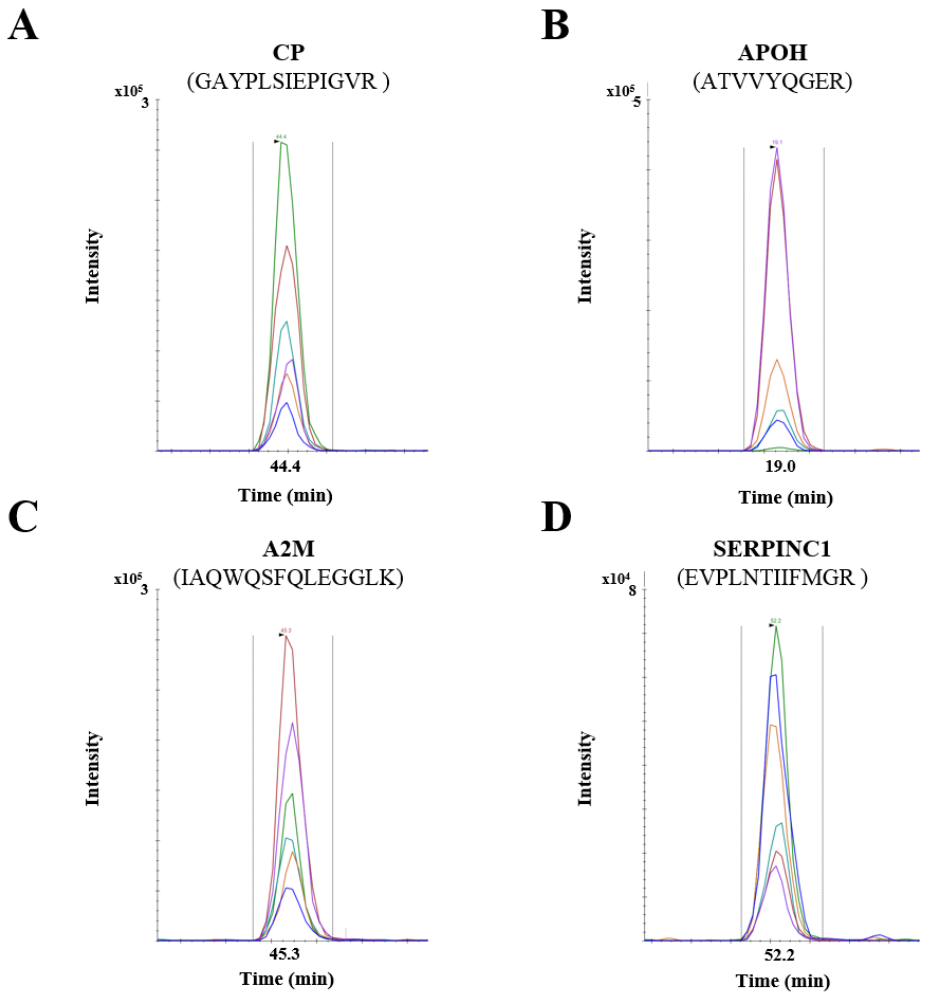

Figure 8. Detectability test of MRM target peptides using QqQ LC-MS.

Four MRM target peptides were analyzed by preliminary MRM analysis with selected top 6 highly abundant MRM transitions obtained from PeptideAtlas DB. Chromatographic elution profiles of selected MRM transitions were generated by unbiased Q3 ion monitoring mode using the QqQ LC-MS

platform. All of the corresponding MRM transitions of (A) GAYPLSIEPIGVR (686.38, MH^{2+}) of CP, (B) ATVVYQGER (511.77, MH^{2+}) of APOH, (C) IAQWQSFQLEGGLK (802.92, MH^{2+}) of A2M, and (D) EVPLNTIIFMGR (695.38, MH^{2+}) of SERPINC1 are assigned with the same retention time.

To improve the quantitative performance of MRM analysis as more accurate and precise manners, we further performed AuDIT analysis using 228 SIS peptides for initial MRM target peptides from preliminary MRM analysis. AuDIT analysis can be used to develop the quantitative MRM method as more accurate and reliable by identifying the inaccurate and inconsistent transitions using SIS peptides (Abbatiello, Mani et al. 2010, Burgess, Keshishian et al. 2014). Through this approach, we removed both interference and poor chromatograms from 228 MRM target peptides with two filter criteria; 1) p-value $< 10^{-5}$ for t-test and 2) CV $\leq$ 20% for PAR. As a result, we determined a total of 135 MRM target peptides derived from 94 target proteins of interest as final MRM targets which could be confidently measured their abundance in our MRM analysis platform (Table 3). Among final MRM target proteins, 10, 44, 55 proteins were derived from public data mining, EOGC profiling, and study sample profiling in our integrative proteomic approach, respectively (Figure 9 and Table 3).

To evaluate biological functions of final MRM target proteins, we used STRING database (version 10.5) to interpret the biological processes of 94 target proteins and confirmed that they were involved in blood coagulation and platelet function, immune response, organelle organization and cellular function categorizations (Figure 10).

Table 3. Final MRM targets by AuDIT analysis.

Uniprot ID	Gene name	Peptide sequence	Target origin	Quantitative MRM transition
P04217	A1BG	HQFLLTGDTQGR	Study sample	686.85/734.34 (y7)
P01023	A2M	LHTEAQIQEEGTVVELTGR	Study sample	704.03/674.38 (y6)
		IAQWQSFQLEGGLK		802.93/978.53 (y9)
P60709	ACTB	DSYVGDEAQSK	Study sample	599.76/734.33 (y7)
Q76LX8	ADAMTS13	FDLELPDGNR	EOGC	588.26/558.26 (y5)
P43652	AFM	IAPQLSTEELVSLGEK	Study sample	857.47/765.41 ($y14^{2+}$)
		ESLLNHFLYEVAR		795.92/897.48 (y7)
P02765	AHSG	HTLNQIDEVK	Study sample	598.82/958.52 (y8)
		FSVVYAK		407.23/579.35 (y5)
P02760	AMBP	TVAACNLPIVR	Public DB/EOGC	607.34/484.32 (y4)
		ETLLQDFR		511.27/565.27 (y4)
P02743	APCS	IVLGQEQDSYGGK	Study sample	697.35/213.16 (b2)
P02647	APOA1	LLDNWDSVTSTFSK	Study sample	806.90/670.34 (y6)
		VSFLSALEEYTK		693.86/940.46 (y8)
P02652	APOA2	SPELQAEAK	Public DB/EOGC	486.75/218.15 (y2)
		EQLTPLIK	/Study sample	471.29/571.38 (y5)
P04114	APOB	FPEVDVLTK	Study sample	524.29/450.76 ($y8^{2+}$)
		TGISPLALIK		506.82/654.45 (y6)
P02654	APOC1	EFGNTLEDK	Public DB/EOGC	526.75/262.14 (y2)
		EWFSETFQK		601.28/739.36 (y6)
P02649	APOE	LGPLVEQGR	EOGC/Study sample	484.78/588.31 (y5)
P02749	APOH	ATVVYQGER	Study sample	511.77/652.30 (y5)
		EHSSLAFWK		552.78/267.11 (b2)
P25311	AZGP1	AGEVQEPELR	Study sample	564.29/514.30 (y4)
		EIPAWVPFDPAAQITK		891.97/770.91 ($y14^{2+}$)
P06276	BCHE	FWTSFFPK	EOGC	530.27/244.17 (y2)
P30043	BLVRB	LQAVTDDHIR	EOGC	389.88/756.36 (y6)
P02746	C1QB	IAFSATR	EOGC/Study sample	388.22/434.24 (y4)
P02747	C1QC	FNAVLTNPQGDYDTSTGK	EOGC/Study sample	964.46/1168.51 (y11)
P09871	C1S	VGATSFYSTCQSNGK	EOGC	803.86/881.38 (y8)

		SSNNPHSPIVEEFQVPYNK		729.36/620.34 (y5)
P01024	C3	SNLDEDIIAEENIVSR	Study sample	908.95/917.47 (y8)
P0C0L4	C4A	ANSFLGEK	Study sample	433.22/680.36 (y6)
		DSSTWLTAFVLK		684.36/791.50 (y7)
P0C0L5	C4B	ASSFLGEK	Study sample	419.72/446.26 (y4)
		VGDTLNLNLR		557.81/629.37 (y5)
P01031	C5	TDAPDLPEENQAR	EOGC/Study sample	728.34/843.40 (y7)
		GIYGTISR		433.74/533.30 (y5)
P10643	C7	VLFYVDSEK	Study sample	550.28/740.35 (y6)
		LSGNVLSYTFQVK		728.40/872.45 (y7)
P07357	C8A	LGSLGAACEQTQTEGAK	Study sample	860.91/991.47 (y9)
		HTSLGPLEAK		526.79/218.15 (y2)
P07360	C8G	QLYGDTGVLGR	Study sample	589.81/774.41 (y8)
		SLPVSDSVLSGFEQR		810.92/710.86 ($y13^{2+}$)
P02748	C9	LSPIYNLVPVK	EOGC/Study sample	621.88/343.23 (y3)
P16070	CD44	YGFIEGHVVIPR	EOGC	462.92/510.30 ($y9^{2+}$)
P12830	CDH1	TIFFCER	EOGC	486.73/758.33 (y5)
P00751	CFB	EAGIPEFYDYDVALIK	Study sample	921.96/736.87 ($y12^{2+}$)
		DFHINLFQVLPWLK		885.49/543.33 (y4)
P08603	CFH	SPDVINGSPISQK	Study sample	671.35/399.19 (b4)
		SSNLIILEEHLK		698.40/768.43 (y6)
P36980	CFHR2	TGDIVEFVCK	Study sample	584.29/274.10 (b3)
P05156	CFI	HGNTDSEGIVEVK	Public DB /Study sample	462.22/246.18 (y2)
		VFSLQWGEVK		596.82/946.50 (y8)
O00533	CHL1	TTVILPLAPFVR	EOGC	663.91/799.48 (y7)
P10909	CLU	ASSIIDELFQDR	Study sample	697.35/922.43 (y7)
P12111	COL6A3	VAVVQYSDR	EOGC	518.77/540.24 (y4)
		ELPSLEQK		472.26/701.38 (y6)
P00450	CP	EYTDASFTNR	Study sample	602.27/624.31 (y5)
		GAYPLSIEPIGVR		686.39/870.50 (y8)
Q96IY4	CPB2	YSFTIELR	Study sample	514.77/631.38 (y5)
P22792	CPN2	LSNNALSGLPQGVFGK	Study sample	801.44/732.40 (y7)
P01034	CST3	ALDFAVGEYNK	EOGC	613.81/780.39 (y7)
Q01459	CTBS	DIIDPAFR	EOGC	473.75/490.28 (y4)

P27487	DPP4	VLEDNSALDK	EOGC	552.28/891.41 (y8)
P17813	ENG	VLPGHSAGPR	EOGC	330.85/487.26 (y5)
P06733	ENO1	IGAEVYHNLK	Public DB	381.88/374.24 (y3)
P00742	F10	TGIVSGFGR	EOGC	447.25/523.26 (y5)
P03951	F11	TAAISGYSFK	EOGC	522.77/688.33 (y6)
		VVSGFSLK		418.75/638.35 (y6)
P00734	F2	ETWTANVGK	EOGC	503.25/488.28 (y5)
P23142	FBLN1	TGYYFDGISR	EOGC	589.78/694.35 (y6)
		ITYYHLSFPTNIQAPAVVFR		779.75/688.41 (y6)
P08637	FCGR3A	AVVFLEPQWYR	EOGC	704.37/749.37 (y5)
P02671	FGA	QLEQVIAK	Public DB	464.78/218.15 (y2)
P02751	FN1	STTPDITGYR	EOGC	555.78/609.34 (y5)
		SYTITGLQPGTDYK		772.39/680.33 (y6)
P02774	GC	VLEPTLK	EOGC/Study sample	400.25/458.30 (y4)
P40197	GP5	YLGVTLSPR	EOGC	503.29/729.43 (y7)
P80108	GPLD1	TLLLVGSPTWK	Study sample	607.86/675.35 (y6)
P22352	GPX3	FLVGPDGIPIMR	Study sample	657.87/516.30 (y4)
P06396	GSN	TGAQELLR	Study sample	444.25/530.33 (y4)
		AGALNSNDAFVLK		660.35/893.47 (y8)
P02790	HPX	NFPSPVDAAFR	Study sample	610.81/480.25 ($y9^{2+}$)
		LYLVQGTQVYVFLTK		886.50/770.44 (y6)
P04196	HRG	DGYLFQLLR	Study sample	562.81/676.41 (y5)
		DSPVLIDFFEDTER		841.90/1058.44 (y8)
P17936	IGFBP3	YGQPLPGYTTK	EOGC	612.82/666.35 (y6)
P19827	ITIH1	AAISGENAGLVR	Study sample	579.32/815.44 (y8)
		FAHYVVTSQVVNTANEAR		669.34/775.37 (y7)
P19823	ITIH2	FYNQVSTPLLR	Study sample	669.36/686.42 (y6)
		FLHVPDTFEGHFDGVPVISK		747.72/872.93 ($y16^{2+}$)
Q06033	ITIH3	SLPEGVANGIEVYSTK	Public DB /Study sample	832.43/732.37 ($y14^{2+}$)
B2RMS9	ITIH4	SPEQQETVLDGNLIIR	Study sample	906.48/685.44 (y6)
		ITFELVYEELLK		748.92/1006.58 (y8)
H0YAC1	KLKB1	IAYGTQGSSGYSLR	Study sample	730.36/638.30 ($y12^{2+}$)
		VLTPDAFVCR		589.31/213.16 (b2)

P01042	KNG1	QVVAGLNFR	Study sample	502.29/677.37 (y6)
		TVGSDTFYSFK		626.30/1051.47 (y9)
P13796	LCP1	VYALPEDLVEVNPK	EOGC	793.43/1139.59 (y10)
P08519	LPA	GSFSTTVTGR	EOGC	506.76/533.30 (y5)
P02750	LRG1	VAAGAFQGLR	Study sample	495.28/748.41 (y7)
		DLLLPQPDLR		590.34/725.39 (y6)
P08253	MMP2	VDAAFNWSK	EOGC	519.26/823.41 (y7)
P26927	MST1	AAFCYQIR	EOGC	514.75/579.32 (y4)
		VVGGHPGNSPWTVSLR		554.96/575.35 (y5)
P13591	NCAM1	FFLCQVAGDAK	EOGC	628.31/848.39 (y8)
Q96S96	PEBP4	VISLLPK	EOGC	385.26/244.17 (y2)
P12955	PEPD	STLFVPR	EOGC	410.24/518.31 (y4)
Q96PD5	PGLYRP2	TFTLLDPK	Study sample	467.77/224.17 (y2)
		TDCPGDALFDLLR		746.86/1116.60 (y10)
P00747	PLG	EAQLPVIENK	Study sample	570.82/699.40 (y6)
P02775	PPBP	TTSGIHPK	Study sample	420.73/638.36 (y6)
		EESLDSDLYAELR		770.36/1081.52 (y9)
P07225	PROS1	NNLELSTPLK	EOGC	564.82/357.25 (y3)
P22891	PROZ	ENFVLTTAK	EOGC	511.78/533.33 (y5)
O00391	QSOX1	DFNIPGFPTVR	EOGC	631.83/472.29 (y4)
Q5VY30	RBP4	FSGTWYAMAK	Study sample	581.27/769.37 (y6)
		YWGVASFLQK		599.82/849.48 (y8)
P06702	S100A9	NIETIINTFHQYSVK	Public DB/EOGC	602.98/618.82 ($y10^{2+}$)
P01011	SERPINA3	ITLLSALVETR	Study sample	608.37/775.43 (y7)
P29622	SERPINA4	LGFTDLFSK	Study sample	514.28/710.37 (y6)
P05543	SERPINA7	NALALFVLPK	Study sample	543.34/244.17 (y2)
P01008	SERPINC1	TSDQIHFFFAK	Study sample	670.84/659.36 (y5)
		EVPLNTIIFMGR		695.38/581.33 ($y10^{2+}$)
P05546	SERPIND1	TLEAQLTPR	Study sample	514.79/685.40 (y6)
		QFPILLDFK		560.82/294.18 (y2)
P36955	SERPINF1	SSFVAPLEK	EOGC/Study sample	489.27/557.33 (y5)
P09486	SPARC	LEAGDHPVELLAR	EOGC	473.92/589.32 ($y11^{2+}$)

Q15582	TGFBI	LTLLAPLNSVFK	EOGC	658.40/988.58 (y9)
P07996	THBS1	SITLFVQEDR	EOGC/Study sample	604.32/793.38 (y6)
P02766	TTR	AADDTWEPFASGK	EOGC	697.81/606.32 (y6)
P04004	VTN	DWHGVPGQVDAAMAGR	EOGC	833.89/595.26 (b5)
		DVWGIEGPIDAAFTR		823.91/947.49 (y9)
L8E853	VWF	ILAGPAGDSNVVK	Study sample	620.85/443.74 ($y9^{2+}$)

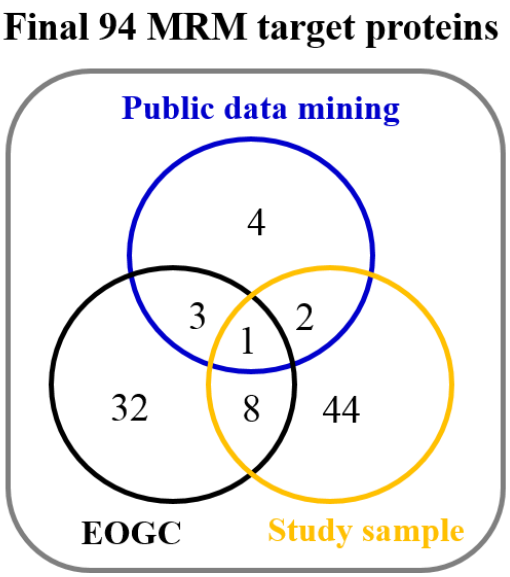

Figure 9. A venn diagram of 94 MRM target proteins obtained from three independent proteomic approaches.

Each number in the cell represents count of MRM target proteins derived from three independent proteomic approaches including public data mining, global proteomic profiling analysis of EOGC patients, and study patient serum samples, respectively.

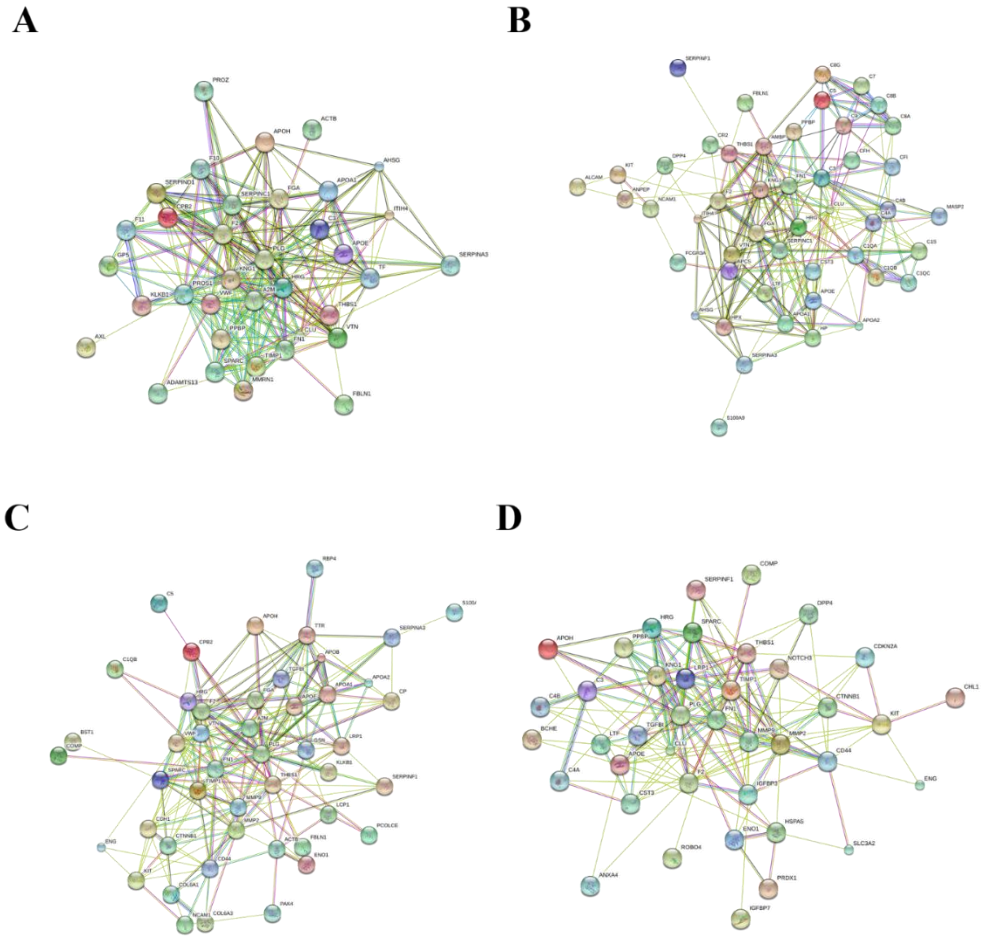

Figure 10. Functional enrichment of MRM target proteins generated from STRING database.

Biological functions of 94 MRM target proteins related to (A) blood coagulation and platelet function, (B) immune response, (C) organelle organization, and (D) cellular function including cell growth, proliferation and apoptosis.

3.5. Multiplexed-MRM analysis of potential serum biomarkers for identification of recurrence in gastric cancers with training set's patient samples

The most reliable strategy for verification of biomarker candidates by clinical proteomic approach is to estimate their abundances with individual samples using MRM analysis. In this study, we measured the relative expression levels of 94 serum biomarker candidates between the GC recurrence group (n=47) and response group (n=133) as a training set. Prior to MRM analysis, we randomized the order of MRM runs with a total of 180 samples to avoid the possible bias derived from technical MS run variations.

We conducted the quantitative MRM analysis with individual training set's patient serum samples (n=180) using the pre-determined 876 MRM transitions for 135 MRM target peptides derived from 94 target proteins of interest with spiked SIS peptides. Figure 11 is a chromatographic trace of 135 MRM peptides in GC patient sample. To obtain the quantitative data of each target peptide from our MRM analysis, we imported quantitative MRM data into Skyline and normalized peak area of target peptides using their corresponding SIS peptides.

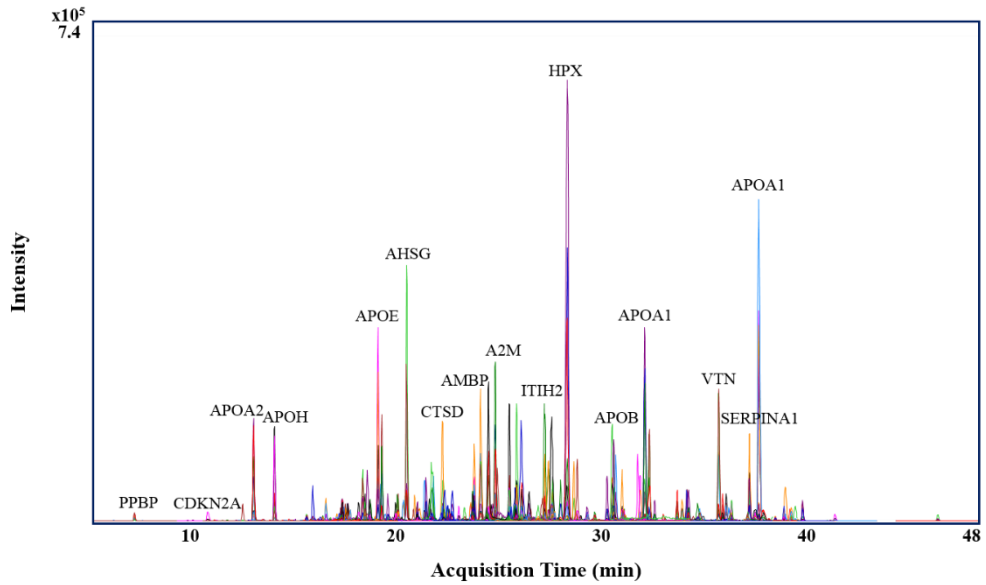

Figure 11. Multiplexed-MRM analysis of potential serum protein biomarkers for identification of recurrence in gastric cancers.

Multiplexed-MRM analysis was performed with 135 peptides of 94 target proteins with their SIS peptides. Extracted ion chromatograms represent the observed 876 MRM transitions 135 MRM target peptides including SIS peptides.

To determine the potential serum biomarkers representing the significant expression levels in the GC recurrence group compared to the response group, we performed statistical analysis with quantitative MRM results by LMM analysis embedded in MSstats. As a result, we identified that 76 target peptides derived from 65 proteins which showed the differentially expressed levels between the two groups with significantly statistical values (p-value < 0.05) (Table 4). We found that the biological functions of these 65 DEPs were involved in blood coagulation, immune response, and cellular proliferation. We further generated HLC with the log_2-transformed peak area's ratio of 76 peptides for 65 DEPs to confirm their distribution patterns in analyzed two groups using Multi Experiment Viewer (Version 4.9) (Figure 12).

Table 4. Significant *DEPs between the two groups by **LMM analysis using MSstats.

Uniprot ID	Gene name	Peptide sequence	logFC	AveExpr	Adjusted *p*-value
P02775	PPBP	TTSGIHPK	-2.34	-0.14	4.45E-19
P27487	DPP4	VLEDNSALDK	-1.62	8.18	1.04E-18
P02652	APOA2	SPELQAEAK	-0.92	-3.94	1.04E-16
P10909	CLU	ASSIIDELFQDR	0.75	-0.61	1.97E-16
P02751	FN1	SYTITGLQPGTDYK	-1.12	-0.19	2.33E-14
P02751	FN1	STTPDITGYR	-1.03	-0.46	8.93E-13
P02750	LRG1	DLLLPQPDLR	0.68	0.72	3.25E-12
P04004	VTN	DWHGVPGQVDAAMAGR	1.31	-0.48	8.14E-12
P01008	SERPINC1	EVPLNTIIFMGR	1.28	-0.45	1.44E-11
L8E853	VWF	ILAGPAGDSNVVK	0.94	5.20	3.56E-11
P26927	MST1	AAFCYQIR	-0.72	1.09	3.62E-10
P02743	APCS	IVLGQEQDSYGGK	-1.78	5.07	1.35E-09
P02671	FGA	QLEQVIAK	1.27	5.52	6.10E-09
O00391	QSOX1	DFNIPGFPTVR	-1.35	1.95	6.74E-09
P22352	GPX3	FLVGPDGIPIMR	0.99	3.40	1.54E-08
P08253	MMP2	VDAAFNWSK	-0.99	7.20	1.88E-08
Q06033	ITIH3	SLPEGVANGIEVYSTK	0.64	1.03	2.15E-08
P02766	TTR	AADDTWEPFASGK	-0.58	-2.38	3.51E-08
P01031	C5	GIYGTISR	0.81	-0.63	4.28E-08
P19823	ITIH2	FLHVPDTFEGHFDGVPVISK	0.44	-2.09	5.47E-08
P06702	S100A9	NIETIINTFHQYSVK	-0.95	1.19	9.41E-08
P01042	KNG1	TVGSDTFYSFK	0.48	-1.95	1.67E-07
P04004	VTN	DVWGIEGPIDAAFTR	0.33	-3.10	2.29E-07
Q5VY30	RBP4	FSGTWYAMAK	1.26	0.62	2.39E-07
P05156	CFI	HGNTDSEGIVEVK	0.35	0.06	2.40E-07
P06276	BCHE	FWTSFFPK	-0.38	0.54	5.56E-07
P00450	CP	GAYPLSIEPIGVR	0.52	-2.70	1.26E-06
P00747	PLG	EAQLPVIENK	0.29	-0.98	1.89E-06
P04196	HRG	DSPVLIDFFEDTER	0.81	-4.52	3.68E-06

P08603	CFH	SPDVINGSPISQK	0.42	-1.06	4.21E-06
P43652	AFM	ESLLNHFLYEVAR	0.38	0.38	6.83E-06
P04114	APOB	TGISPLALIK	-0.43	-1.09	7.94E-06
P36955	SERPINF1	SSFVAPLEK	-0.32	1.81	1.41E-05
P02790	HPX	NFPSPVDAAFR	-0.38	-4.47	2.84E-05
P02746	C1QB	IAFSATR	-1.02	1.60	3.06E-05
P22891	PROZ	VVSGFSLK	0.48	5.32	3.22E-05
P00734	F2	ETWTANVGK	0.57	0.67	4.34E-05
P02654	APOC1	EWFSETFQK	0.50	-0.16	5.48E-05
P02765	AHSG	HTLNQIDEVK	0.80	-1.56	6.57E-05
P02654	APOC1	EFGNTLEDK	0.75	1.48	8.06E-05
Q15582	TGFBI	LTLLAPLNSVFK	0.34	4.57	8.17E-05
P00751	CFB	EAGIPEFYDYDVALIK	0.51	-2.17	0.0002
P04114	APOB	FPEVDVLTK	-0.50	2.26	0.0002
P09486	SPARC	LEAGDHPVELLAR	0.81	5.64	0.0002
P30043	BLVRB	LQAVTDDHIR	-0.50	4.20	0.0003
P05546	SERPIND1	QFPILLDFK	-0.32	-0.72	0.0004
P04217	A1BG	HQFLLTGDTQGR	0.25	-2.24	0.0004
P23142	FBLN1	TGYYFDGISR	-0.32	1.49	0.0005
P07357	C8A	HTSLGPLEAK	0.25	1.30	0.0005
P25311	AZGP1	EIPAWVPFDPAAQITK	0.34	0.40	0.0007
P02774	GC	VLEPTLK	0.28	-2.89	0.0007
P07357	C8A	LGSLGAACEQTQTEGAK	0.36	0.79	0.0015
P07996	THBS1	SITLFVQEDR	-0.46	-0.25	0.0015
P03951	F11	TAAISGYSFK	0.29	4.31	0.0017
P09871	C1S	SSNNPHSPIVEEFQVPYNK	-0.40	-0.61	0.0020
P09871	C1S	VGATSFYSTCQSNGK	0.45	-0.12	0.0029
P12955	PEPD	STLFVPR	-0.24	5.37	0.0030
P02765	AHSG	FSVVYAK	0.24	-2.96	0.0038
P16070	CD44	YGFIEGHVVIPR	0.35	2.99	0.0039
P29622	SERPINA4	LGFTDLFSK	-0.16	1.60	0.0059
P13591	NCAM1	FFLCQVAGDAK	0.34	5.35	0.0070
Q96IY4	CPB2	YSFTIELR	-0.31	1.77	0.0084
P01023	A2M	IAQWQSFQLEGGLK	0.26	-3.89	0.0104

P00450	CP	EYTDASFTNR	0.22	-0.78	0.0110
P22891	PROZ	ENFVLTTAK	-0.28	4.35	0.0110
P02652	APOA2	EQLTPLIK	-0.42	-2.63	0.0118
P02749	APOH	EHSSLAFWK	0.41	-1.90	0.0123
B2RMS9	ITIH4	SPEQQETVLDGNLIIR	-0.25	-2.04	0.0131
P06396	GSN	AGALNSNDAFVLK	0.21	-0.29	0.0158
P60709	ACTB	DSYVGDEAQSK	-0.32	5.11	0.0160
B2RMS9	ITIH4	ITFELVYEELLK	-0.51	-2.32	0.0161
P01011	SERPINA3	ITLLSALVETR	-0.21	-3.52	0.0163
P17813	ENG	VLPGHSAGPR	0.14	7.49	0.0202
P12830	CDH1	TIFFCER	0.22	3.48	0.0233
O00533	CHL1	TTVILPLAPFVR	0.17	5.76	0.0285
P08519	LPA	GSFSTTVTGR	0.67	5.30	0.0308

* DEPs : Differentially expressed proteins generated by MSstats analysis.

**LMM : Linear mixed model.

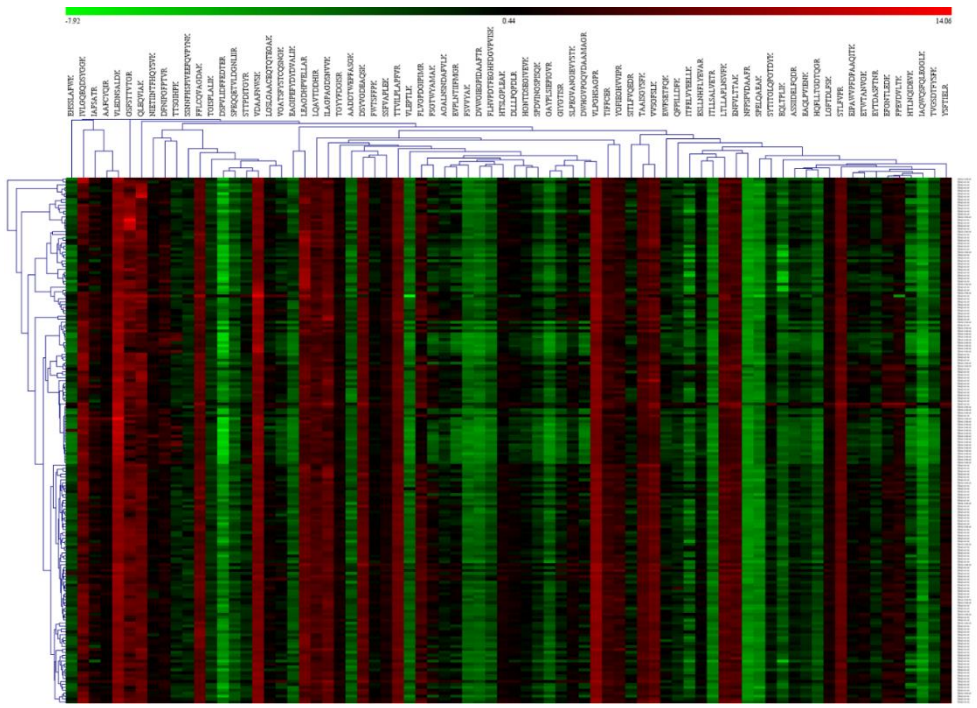

Figure 12. Cluster analysis with 65 DEPs in GC response versus recurrence groups.

HLC analysis was conducted with 76 peptides for 65 DEPs (p-value < 0.05) in a training set using Multi Experiment Viewer (Version 4.9). The x- and y-axes represent each target peptide and individual samples, respectively.

3.6. Construction of the multi-biomarker panel for identification of GC response and recurrence groups

In this study, our final goal was to construct the multi-biomarker panel which could predict GC patients who have the potential possibility of recurrence events. Although 65 proteins showed significant changes in their abundances between the two groups, these proteins had the limitations as a single marker for identification of GC recurrence with regard to their specificity and sensitivity. Therefore, we constructed the multi-biomarker panel by combining 65 DEPs to improve the discriminatory power for identification of GC recurrence and response groups. We performed logistic regression analysis using the stepwise method and generated the multi-biomarker panel's equation with 6 proteins including APOA2, APOC1, CLU, ITIH4, LRG1, and SERPINA3, which represented the best combination derived from 65 DEPs. Among 6 marker proteins, APOA2, APOC1, and other 4 proteins (CLU, ITIH4, LRG1, and SERPINA3) were originated from all three integrative proteomic approaches, public data mining and EOGC profiling, and study sample profiling, respectively. The molecular function of these 6 proteins were involved in enzyme inhibitor activity, phosphatidylcholine binding, and lipase inhibitor activity, associated with cancer recurrence (Hudler, Kocevar et al. 2014, Vargas, Moreno-Rubio et al. 2014, Murakami, Kubo et al. 2015). Although each protein of 6-marker panel showed a little lower AUC values

(AUC=0.550, 0.732, 0.757, 0.512, 0.771, and 0.570 in APOA2, APOC1, CLU, ITIH4, LRG1, and SERPINA3, respectively) (Figure 13) as a single marker in analyzed patient samples, we found that when the combination of these proteins as the multi-biomarker panel, which showed a higher AUC value exceeded 0.8 (AUC = 0.810) (Figure 14). More importantly, the 6-marker panel also showed a great prediction rate for distinguishing the two groups, which could identify 127 of 133 (95.5%) in the response group and 29 of 47 (61.7%) in the recurrence group (Table 5).

In addition, we measured the absolute concentration of 6 marker proteins in the two groups as in known amounts of their SIS peptides from quantitative MRM data as shown in Table 6.

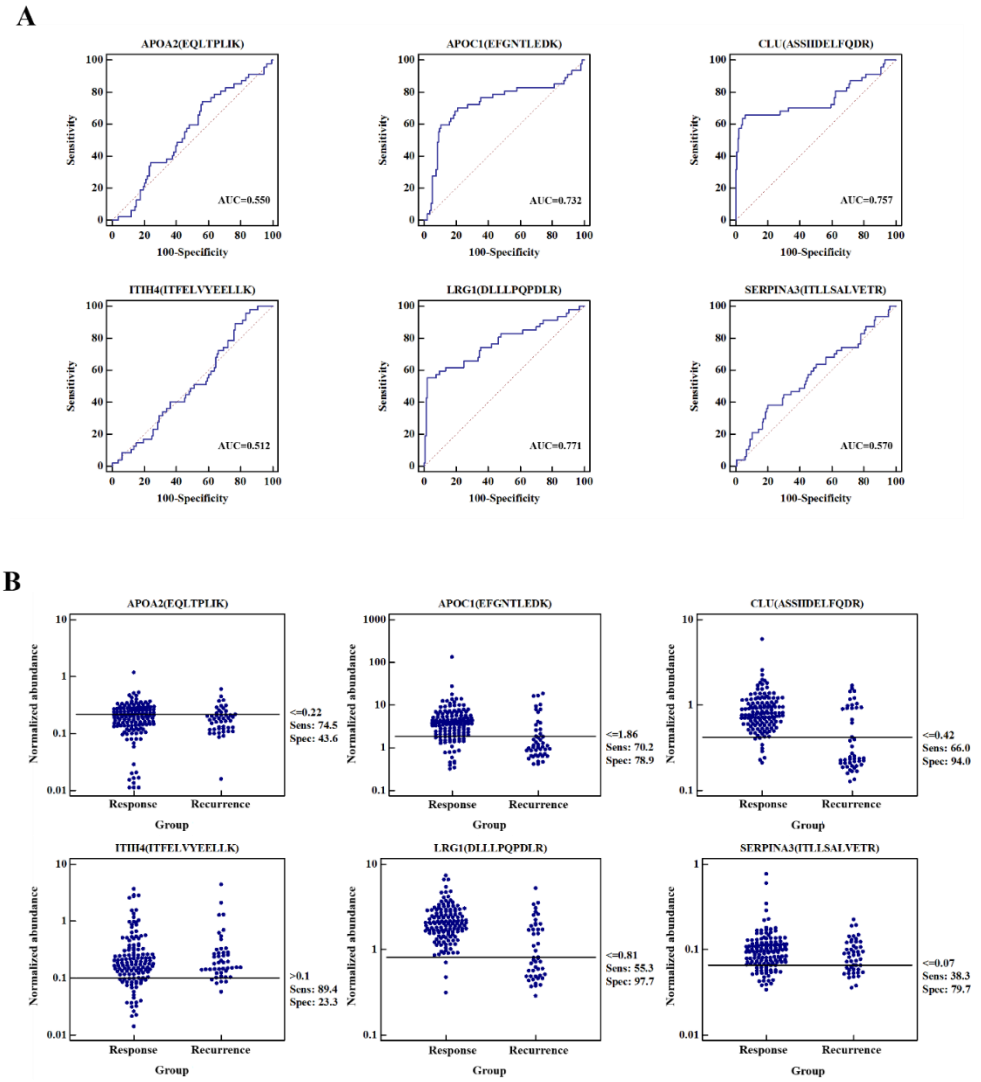

Figure 13. ROC curves and interactive plots for independent six marker proteins in a training set.

The normalized peak areas of transitions were compared between GC response (n=133) and recurrence (n=47) groups. (A) The AUC value of each protein

showed 0.550, 0.732, 0.757, 0.512, 0.771, and 0.570 in APOA2, APOC1, CLU, ITIH4, LRG1, and SERPINA3, respectively. (B) The interactive plots were represented by the log_2-transformed peak area's ratio (endogenous target peptide versus SIS peptide) of 6 proteins with regard to relative concentration, sensitivity, and specificity. All data was generated from MedCalc software (Version 12.2.1.).

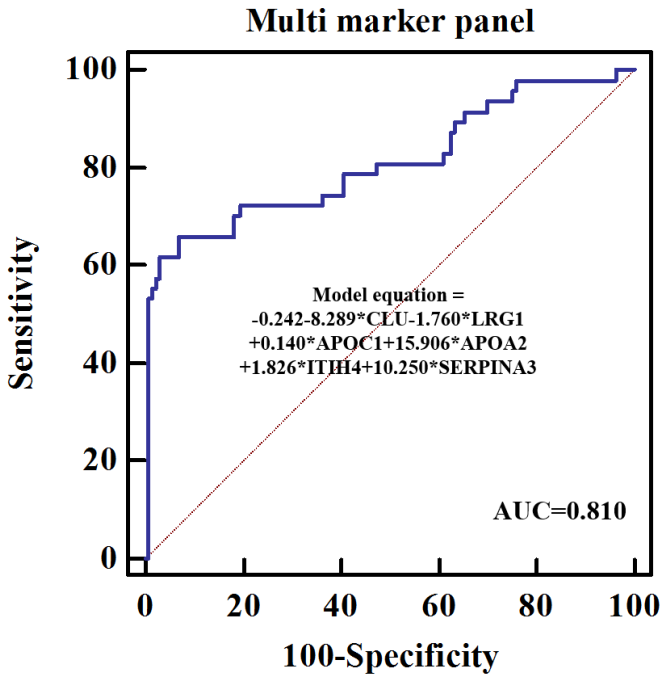

Figure 14. A ROC curve of the constructed 6-marker panel in a training set.

Logistic regression analysis was performed with a combined 6-marker panel's equation value using the stepwise method to generate ROC curve. The AUC value of 6-marker panel showed 0.810 which was higher than each protein's AUC value as described in Figure 13.

Table 5. Prediction rate of the 6-marker panel for discriminating the two groups.

Predicted group / Actual group	Response group	Recurrence group	Prediction rate
Response group	127	6	95.49%
Recurrence group	18	29	61.70%

Sensitivity = 66.0, Specificity = 93.2

Table 6. Absolute concentration of 6 marker proteins in the two groups.

6-marker panel	Average concentration (㎍/㎖)	
	Response group	Recurrence group
APOA2	238.4	257.7
APOC1	9.1	16.3
CLU	54.6	96.2
ITIH4	133.5	129.5
LRG1	22.3	38.8
SERPINA3	461.0	546.2

3.7. Verification of the 6-marker panel's quantitative performance in MRM analysis

To demonstrate the reasonable quantitative performance of the 6-marker panel in complex serum matrices by MRM analysis, we generated the quantitative linearity curve using their corresponding 6 SIS peptides. This was accomplished by measuring the MRM responses at 10 different concentrations (1, 2.5, 5, 10, 25, 50, 100, 250, 500, 1000 fmol) of 6 SIS peptides spiked into the pooled GC patient serum samples. The quantitative curve resulted in good linearity values (R^2 = 0.9998, 0.9997, 0.9878, 0.9705, 0.9994, and 0.9951 in EQLTPLIK of APOA2, EFGNTLEDK of APOC1, ASSIIDELFQDR of CLU, ITFELVYEELLK of ITIH4, DLLLPQPDLR of LRG1, and ITLLSALVETR of SERPINA3, respectively) with reproducible CV values in triplicate MRM analysis (Figure 15). These results demonstrated that our 6-protein marker panel might be measured their abundances with a highly robust and reproducible quantitative manner in complex serum matrices by MRM analysis.

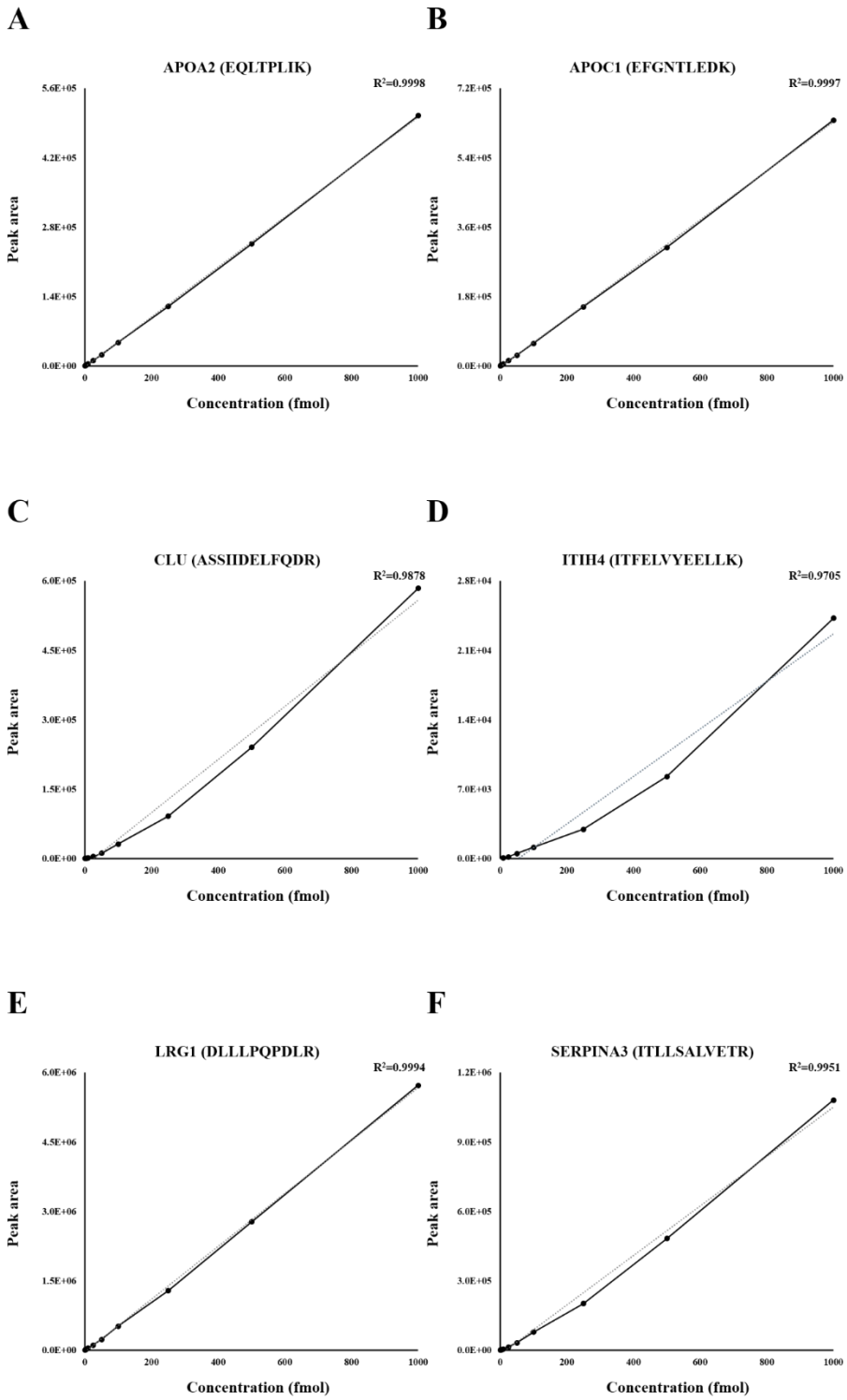
A
APOA2 (EQLTPLIK)
R²=0.9998
Peak area
Concentration (fmol)
B
APOC1 (EFGNTLEDK)
R²=0.9997
Peak area
Concentration (fmol)
C
CLU (ASSIIDELFQDR)
R²=0.9878
Peak area
Concentration (fmol)
D
ITIH4 (ITFELVYEELLK)
R²=0.9705
Peak area
Concentration (fmol)
E
LRG1 (DLLLPQPDLR)
R²=0.9994
Peak area
Concentration (fmol)
F
SERPINA3 (ITLLSALVETR)
R²=0.9951
Peak area
Concentration (fmol)

Figure 15. Quantitative linearity curves of the 6-protein marker panel using their corresponding SIS peptides.

MRM assay was performed in triplicate at 10 concentrations (1, 2.5, 5, 10, 25, 50, 100, 250, 500, and 1000 fmol) of 6 SIS peptides (EQLTPLIK, EFGNTLEDK, ASSIIDELFQDR, ITFELVYEELLK, DLLLPQPDLR, and ITLLSALVETR of APOA2, APOC1, CLU, ITIH4, LRG1, and SERPINA3, respectively) spiked in GC patient serum samples. The quantitative curve resulted in a linearity of R^2=0.9998, 0.9997, 0.9878, 0.9705, 0.9994, and 0.9951 in (A) APOA2, (B) APOC1, and (C) CLU, (D) ITIH4, (E) LRG1, and (F) SERPINA3, respectively.

3.8. Verification of the 6-marker panel with test set's patient samples by MRM analysis

To verify the performance of the 6-marker panel, we applied multiplexed-MRM analysis with independent GC patient samples (n=64), i.e. response group (n=43) and recurrence group (n=21), as a test set. In this approach, we used the same analyzed conditions such as QqQ LC-MS parameters, MRM targets, and randomization of sample analysis order as previously applied in the training set.

With the generated MRM quantitative data from individual test set's serum samples, we performed MSstats analysis whether the 6-marker panel had the reproducible performance as in training set's results. As a result, we observed that the 6-marker panel had the correlated expressed patterns compared to training set's results with significantly statistical values (p-value < 0.05) (Table 7) suggesting that the 6-marker panel exhibited consistent quantitative features in both training and test sets.

Using all 244 patient's quantitative MRM data, we also evaluated both AUC value and prediction rate of the 6-marker panel based on DFS at 5 years. Among a total of 244 GC patients, 169 patients (69.3%) were alive and other 75 patients (30.7%) were expired at 5 years. Although 6-marker panel did not show the significant AUC value (AUC=0.699), the 6-marker panel showed a

reasonable prediction rate of 5-year DFS, which could identify 160 of 169 (94.7%) in alive patients and 29 of 75 (38.7%) in expired patients at 5 years.

Table 7. Comparison of the 6-marker panel between training and test sets.

Uniprot ID	Gene name	Peptide sequence	Training set		Test set	
			AveExpr	Adjusted *p*-value	AveExpr	Adjusted *p*-value
P02652	APOA2	EQLTPLIK	-2.63	0.0117	-2.24	0.0378
P02654	APOC1	EFGNTLEDK	1.48	8.06E-05	2.07	0.0317
P10909	CLU	ASSIIDELFQDR	-0.61	1.97E-16	-0.12	0.0313
Q14624	ITIH4	ITFELVYEELLK	-2.32	0.0161	-4.40	0.0019
P02750	LRG1	DLLLPQPDLR	0.72	3.25E-12	1.05	0.0406
P01011	SERPINA3	ITLLSALVETR	-3.52	0.0161	-3.55	0.0311

4. Discussion

In the course of biomarker development for identification of GC recurrence patients who received chemotherapy after D2 lymph node dissection, we discovered and verified the multi-biomarker panel by quantitative proteomic approach using MRM analysis. In the discovery stage, we selected 330 initial biomarker candidate proteins by integrative proteomic approach using three independent strategies including public DB mining, EOGC profiling, and study sample profiling. We performed preliminary MRM analysis with 330 initial biomarker candidates to confirm their detectability in QqQ LC-MS system and selected 228 peptides derived from 141 proteins as first MRM targets. Most importantly, we then conducted AuDIT analysis with 228 MRM target peptides using their corresponding SIS peptides to determine final MRM targets which may have the accurately and reliably quantitative manners in MRM analysis. Through these approaches, a set of 135 peptides derived from 94 proteins were selected final MRM targets as biomarker candidates for identification of GC recurrence. In general, at least two ionisable unique peptides and their two or three transitions could be chosen from the tryptic digest of the protein of interest for quantification using MRM analysis. In this study, 53 proteins were estimated their abundances with one unique target peptide, however, we used more than 4 transitions (4~8) for each target peptide

to quantify their abundance confidently. For these reasons, we could estimate the accurate quantification of 94 target proteins from our MRM analysis.

In validation stage, we performed the measurement of differentially expressed levels of MRM target proteins with 180 individual patient samples, GC recurrence group (n=47) and response group (n=133), as a training set. We applied a stringent statistical analysis using MRM quantitative data from individual samples by MSstats analysis and identified that 65 proteins showed differentially expressed levels with significant values (p-value < 0.05) between the two groups. However, these proteins have some limitations as a single marker for identification of GC recurrence due to their specificity and sensitivity. Therefore, we further conducted a logistic regression analysis to generate the multi-biomarker panel by combination of these 65 DEPs. As a result, we identified the 6-marker panel, comprising of APOA2, APOC1, CLU, ITIH4, LRG1, and SERPINA3, which represented a significant AUC value (0.810) and high prediction rates for distinguishing two groups (95.5% and 61.7% in response and recurrence groups, respectively).

Although the detailed mechanism is less well-known, our 6-protein marker panel may be related to GC and/or several cancer recurrence mechanisms. Recently, several studies have indicated that apolipoproteins including APOA2 and APOC1 are mainly synthesized in the liver and may be involved

in GC biology mechanisms triggering programmed cell death (Vanhollebeke and Pays 2006, Liu, Pan et al. 2012). Especially APOC1 has been considered potential serum biomarker for GC by discriminating GC patients from cancer-free controls (Cohen, Yossef et al. 2011). APOC1 binds and enhances the inflammatory response which promotes tumor development and progression and maintains cell survival by anti-apoptotic and proliferation-enhancing effects on cancer cells. Also, APOC1 has an inverse relationship with several cancers, such as colorectal, breast, lung, and pancreatic cancers (Engwegen, Helgason et al. 2006, Takano, Yoshitomi et al. 2008, Fan, Wang et al. 2010, Xue, Scarlett et al. 2010).

CLU is a ubiquitous secretory sulfated glycoprotein and is an inhibitor of apoptosis with a cytoprotective function (Zhang, Kim et al. 2005). The expression level of CLU is associated with tumor progression as anti-apoptotic mediator in bladder, breast, gastric, pancreatic, and prostate cancers (Miyake, Hara et al. 2001, Pins, Fiadjoe et al. 2004, Bi, Guo et al. 2009, Yom, Woo et al. 2009, Jin, Kim et al. 2012). In the patients with <T2 stage breast cancer, CLU overexpression was a significant prognostic factor for recurrence in spite of negative lymph node status and postoperative systemic treatment in all cases (Yom, Woo et al. 2009). More importantly, overexpression of CLU was associated with tumor invasion and metastasis as p53 abnormal expression in

GC patients who showed significantly poor prognosis in late stage (Bi, Guo et al. 2009).

ITIH4 is a member of the inter-alpha-trypsin inhibitor (ITI) family and an acute-phase reactant elevated in response to interleukin-6 (Piñeiro, Alava et al. 1999). It is well known that ITIH genes are potential candidates as tumor suppressor or metastasis repressor genes (Fries and Blom 2000, Kobayashi, Suzuki et al. 2004). Thus, ITIH molecules may serve as potential diagnostic markers or therapeutic targets in the malignant setting. In particular, it is known that ITIH4 is observed with elevated levels in GC patients associating with tumor progression, invasion and metastasis with increased proliferation rates (Subbannayya, Mir et al. 2015).

LRG1 is an oncogene-associated protein which has been clarified vital to various cancers such as gastric, leukemia, non-small cell lung, ovarian, and pancreatic cancers (Kakisaka, Kondo et al. 2007, Andersen, Boylan et al. 2010, Liu, Luo et al. 2012, Wu, Yu et al. 2012, Uen, Lin et al. 2013). The key function of LRG1 in cancer states is cellular survival and apoptosis activating TFG-beta signaling pathway (Wang, Abraham et al. 2013). Also, LRG1 has been previously reported as overexpressed levels in GC patient plasma and/or serum samples as associating to promote angiogenesis (Chong, Lee et al. 2010).

The aberrant expression level of SERPINA3 has been observed in various human cancers (Cimino, Fuso et al. 2008, Pascal, Vêncio et al. 2009, Gelfand, Vernet et al. 2017). SERPINA3 is a member of the serpin super-family of protease inhibitors and usually shown up-regulated in cancer patients by regulating both cell cycles and apoptosis via activation of MAPK/ERK and PI3K/AKT signaling pathways (Yang, Yang et al. 2014).

Although functional studies are needed to evaluate the 6-marker panel further, our MRM results suggested that the 6 marker proteins, which showed the differentially expressed levels in GC response and recurrence groups, may play important roles in tumor metastasis, invasion, and apoptotic processes associated with GC recurrence mechanisms as previously reported.

Finally, we applied multiplexed-MRM analysis with independent 64 GC patient samples, GC recurrence group (n=43) and response group (n=21), to verify the 6-marker panel whether it shows the similar quantitative features as in the training set. As a result, we found that the 6-marker panel showed the correlated expression patterns in test set's results with significant statistical values (p-value < 0.05). In addition, we performed logistic regression analysis to measure the 6-marekr panel's prediction rates in the test set and confirmed that this panel could identify 43 of 43 (100.0%) in response group and 0 of 21 (0.0%) in recurrence group. Although 6-marker panel showed reproducibly

expressed patterns in both training and test sets, it could not predict recurrence group in the test set. We thought that this negative result may arise from few number of recurrence individual samples to obtain the confident statistical values in the test set. However, we speculated that these results also indicated the positive aspect for GC patients who should select the potential chemotherapy treatments after D2 lymph node dissection, because the 6-marker panel could identify all response group's patients.

There are two major limitations in this study. First, although all study serum samples were collected, stored, and processed using predefined standard operating procedures, we obtained these samples from a single site clinical center (Samsung Medical Center, Seoul, Korea). Ideally, the protein biomarker validation study should be performed with independent patient cohorts collected from multiple clinical centers. Next, we carried out multiplexed-MRM analysis with limited number of individual patient samples (n=244). Especially the ratio of controls (response group, n=176) and matched cases (recurrence group, n=68) were restricted to ensure adequate statistical power in this study. Based on these limitations, we focused on the technical aspects of quantitative MRM assay as a robust biomarker development platform by measuring biomarker candidate's abundances confidently. As a result, 6 proteins as generated the multi-marker panel achieved CV values less than 20%

with an adequate diagnostic AUC value (AUC=0.810).

In conclusion, we established a quantitative MRM-based serum biomarker development platform as high-throughput and reproducible assay which could serve as a valuable tool in the biomarker discovery-validation process. Through our approach, we could construct the multi-biomarker panel for identification of GC recurrence patients. Although further studies should be required to validate the 6-marker panel using independent large cohorts ($n \geq 500$) collected from multiple hospitals, we propose that our data may contribute to identify the GC response and recurrence patients who treated chemotherapy after D2 lymph node dissection by prediction of their recurrence possibility.

5. References

Abbatiello, S. E., D. Mani, H. Keshishian and S. A. Carr (2010). "Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry." Clinical chemistry **56**(2): 291-305.

Andersen, J. D., K. L. Boylan, R. Jemmerson, M. A. Geller, B. Misemer, K. M. Harrington, S. Weivoda, B. A. Witthuhn, P. Argenta and R. I. Vogel (2010). "Leucine-rich alpha-2-glycoprotein-1 is upregulated in sera and tumors of ovarian cancer patients." Journal of ovarian research **3**(1): 21.

Bi, J., A. Guo, Y. Lai, B. Li, J. Zhong, H. Wu, Z. Xie, Y. He, Z. Lv and S. Lau (2009). "Overexpression of clusterin correlates with tumor progression, metastasis in gastric cancer: a study on tissue microarrays." Neoplasma **57**(3): 191-197.

Burgess, M. W., H. Keshishian, D. Mani, M. A. Gillette and S. A. Carr (2014). "Simplified and efficient quantification of low-abundance proteins at very high multiplex via targeted mass spectrometry." Molecular & Cellular Proteomics **13**(4): 1137-1149.

Carcas, L. P. (2014). "Gastric cancer review." Journal of carcinogenesis **13**(1): 14.

Cho, B. K., Y. D. Koo, K. Kim, M. J. Kang, Y. Y. Lee, Y. Kim, K. S. Park, K.

P. Kim and E. C. Yi (2014). "Determination of selected reaction monitoring peptide transitions via multiplexed product-ion scan modes." Rapid Communications in Mass Spectrometry **28**(7): 773-780.

Choi, M., C.-Y. Chang, T. Clough, D. Broudy, T. Killeen, B. MacLean and O. Vitek (2014). "MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments." Bioinformatics: btu305.

Chong, P. K., H. Lee, M. C. S. Loh, L. Y. Choong, Q. Lin, J. B. Y. So, K. H. Lim, R. A. Soo, W. P. Yong and S. P. Chan (2010). "Upregulation of plasma C9 protein in gastric cancer patients." Proteomics **10**(18): 3210-3221.

Cimino, D., L. Fuso, C. Sfiligoi, N. Biglia, R. Ponzone, F. Maggiorotto, G. Russo, L. Cicatiello, A. Weisz and D. Taverna (2008). "Identification of new genes associated with breast cancer progression by gene expression analysis of predefined sets of neoplastic tissues." International journal of cancer **123**(6): 1327-1338.

Cohen, M., R. Yossef, T. Erez, A. Kugel, M. Welt, M. M. Karpasas, J. Bones, P. M. Rudd, J. Taieb and H. Boissin (2011). "Serum apolipoproteins CI and C-III are reduced in stomach cancer patients: results from MALDI-based peptidome and immuno-based clinical assays." PloS one **6**(1): e14540.

D'Angelica, M., M. Gonen, M. F. Brennan, A. D. Turnbull, M. Bains and M. S. Karpeh (2004). "Patterns of initial recurrence in completely resected gastric adenocarcinoma." Annals of surgery **240**(5): 808-816.

DeLong, E. R., D. M. DeLong and D. L. Clarke-Pearson (1988). "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach." Biometrics: 837-845.

Ebert, M. P. and C. Röcken (2006). "Molecular screening of gastric cancer by proteome analysis." European journal of gastroenterology & hepatology **18**(8): 847-853.

Elimova, E., R. Wadhwa, H. Shiozaki, K. Sudo, J. S. Estrella, B. D. Badgwell, P. Das, A. Matamoros, S. Song and J. A. Ajani (2015). "Molecular biomarkers in gastric cancer." Journal of the National Comprehensive Cancer Network **13**(4): e19-e29.

Engwegen, J. Y., H. H. Helgason, A. Cats, N. Harris, J. M. Bonfrer, J. H. Schellens and J. H. Beijnen (2006). "Identification of serum proteins discriminating colorectal cancer patients and healthy controls using surface-enhanced laser desorption ionisation-time of flight mass spectrometry." World journal of gastroenterology: WJG **12**(10): 1536.

Fan, Y., J. Wang, Y. Yang, Q. Liu, Y. Fan, J. Yu, S. Zheng, M. Li and J. Wang (2010). "Detection and identification of potential biomarkers of breast cancer." Journal of cancer research and clinical oncology **136**(8): 1243-1254.

Ferro, A., B. Peleteiro, M. Malvezzi, C. Bosetti, P. Bertuccio, F. Levi, E. Negri, C. La Vecchia and N. Lunet (2014). "Worldwide trends in gastric cancer mortality (1980–2011), with predictions to 2015, and incidence by subtype."

European Journal of Cancer **50**(7): 1330-1344.

Frantzi, M., A. Bhat and A. Latosinska (2014). "Clinical proteomic biomarkers: relevant issues on study design & technical considerations in biomarker development." Clinical and translational medicine **3**(1): 1.

Fries, E. and A. M. Blom (2000). "Bikunin—not just a plasma proteinase inhibitor." The international journal of biochemistry & cell biology **32**(2): 125-137.

Gelfand, R., D. Vernet, K. W. Bruhn, S. Sarkissyan, D. Heber, J. V. Vadgama and N. F. Gonzalez-Cadavid (2017). "Long-term exposure of MCF-7 breast cancer cells to ethanol stimulates oncogenic features." International Journal of Oncology **50**(1): 49-65.

Gerber, S. A., J. Rush, O. Stemman, M. W. Kirschner and S. P. Gygi (2003). "Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS." Proceedings of the National Academy of Sciences **100**(12): 6940-6945.

Griss, J., Y. Perez-Riverol, H. Hermjakob and J. A. Vizcaino (2015). "Identifying novel biomarkers through data mining—A realistic scenario?" PROTEOMICS-Clinical Applications **9**(3-4): 437-443.

Haruma, K., K. Komoto, T. Kamada, M. Ito, Y. Kitadai, M. Yoshihara, K. Sumii and G. Kajiyama (2000). "Helicobacter pylori infection is a major risk factor for gastric carcinoma in young patients." Scandinavian journal of

gastroenterology **35**(3): 255-259.

Holian, O., S. Wahid, M. J. Atten and B. M. Attar (2002). "Inhibition of gastric cancer cell proliferation by resveratrol: role of nitric oxide." American Journal of Physiology-Gastrointestinal and Liver Physiology **282**(5): G809-G816.

Hudler, P., N. Kocevar and R. Komel (2014). "Proteomic approaches in biomarker discovery: new perspectives in cancer diagnostics." The scientific world journal **2014**.

Humphries, J. M., M. A. Penno, F. Weiland, M. Klingler-Hoffmann, A. Zuber, A. Boussioutas, M. Ernst and P. Hoffmann (2014). "Identification and validation of novel candidate protein biomarkers for the detection of human gastric cancer." Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics **1844**(5): 1051-1058.

Jin, J., J.-M. Kim, Y.-S. Hur, W. P. Cho, K.-Y. Lee, S.-I. Ahn, K. C. Hong and I.-S. Park (2012). "Clinical significance of clusterin expression in pancreatic adenocarcinoma." World journal of surgical oncology **10**(1): 146.

Jin, Z., W. Jiang and L. Wang (2015). "Biomarkers for gastric cancer: Progression in early diagnosis and prognosis (Review)." Oncology letters **9**(4): 1502-1508.

Kakisaka, T., T. Kondo, T. Okano, K. Fujii, K. Honda, M. Endo, A. Tsuchida, T. Aoki, T. Itoi and F. Moriyasu (2007). "Plasma proteomics of pancreatic cancer patients by multi-dimensional liquid chromatography and two-

dimensional difference gel electrophoresis (2D-DIGE): up-regulation of leucine-rich alpha-2-glycoprotein in pancreatic cancer." Journal of Chromatography B **852**(1): 257-267.

Killcoyne, S., E. W Deutsch and J. Boyle (2012). "Mining peptideatlas for biomarkers and therapeutics in human disease." Current pharmaceutical design **18**(6): 748-754.

Kim, D. J., J. H. Lee and W. Kim (2014). "Very early-onset peritoneal recurrence following curative total gastrectomy for Borrmann 4 gastric cancer." Annals of surgical treatment and research **86**(1): 45-49.

Kirkpatrick, D. S., N. A. Hathaway, J. Hanna, S. Elsasser, J. Rush, D. Finley, R. W. King and S. P. Gygi (2006). "Quantitative analysis of in vitro ubiquitinated cyclin B1 reveals complex chain topology." Nature cell biology **8**(7): 700-710.

Klecka, W. R. (1980). Discriminant analysis, Sage.

Kobayashi, H., M. Suzuki, N. Kanayama and T. Terao (2004). "Genetic down-regulation of phosphoinositide 3-kinase by bikunin correlates with suppression of invasion and metastasis in human ovarian cancer HRA cells." Journal of Biological Chemistry **279**(8): 6371-6379.

Lastraioli, E., M. Raffaella Romoli and A. Arcangeli (2012). "Immunohistochemical biomarkers in gastric cancer research and management." International journal of surgical oncology **2012**.

Leal, M. F., P. P. Assumpção, M. C. Smith and R. R. Burbano (2014). "Searching for gastric cancer biomarkers through proteomic approaches." Journal of Gastroenterology and Hepatology Research **3**(3): 989-995.

Li, G.-C., Z. Zhang, X.-J. Ma, W.-L. Gu, Y.-N. Wang and J. Li (2012). "Are biomarkers correlated with recurrence patterns in patients with resectable gastric adenocarcinoma." Molecular biology reports **39**(1): 399-405.

Lin, L.-L., H.-C. Huang and H.-F. Juan (2012). "Discovery of biomarkers for gastric cancer: a proteomics approach." Journal of proteomics **75**(11): 3081-3097.

Liu, C., C. Pan and Y. Liang (2012). "Screening and identification of serum proteomic biomarkers for gastric adenocarcinoma." Experimental and therapeutic medicine **3**(6): 1005-1009.

Liu, Y., X. Luo, H. Hu, R. Wang, Y. Sun, R. Zeng and H. Chen (2012). "Integrative proteomics and tissue microarray profiling indicate the association between overexpressed serum proteins and non-small cell lung cancer." PloS one **7**(12): e51748.

Maehara, Y., Y. Emi, S. Tomisaki, T. Oshiro, Y. Kakeji, Y. Ichiyoshi and K. Sugimachi (1996). "Age-related characteristics of gastric carcinoma in young and elderly patients." Cancer **77**(9): 1774-1780.

McDonald, W. H. and J. R. Yates (2002). "Shotgun proteomics and biomarker discovery." Disease markers **18**(2): 99-105.

Mi, L., X. Ji and J. Ji (2016). "Prognostic biomarker in advanced gastric cancer." Translational Gastrointestinal Cancer **5**(1): 16-29.

Milne, A. N. and G. J. A. Offerhaus (2010). "Early-onset gastric cancer: Learning lessons from the young." World Journal of Gastrointestinal Oncology [E] **2**(2): 59-64.

Milne, A. N., R. Sitarz, R. Carvalho, F. Carneiro, A. Offerhaus and G. Johan (2007). "Early onset gastric cancer: on the road to unraveling gastric carcinogenesis." Current molecular medicine **7**(1): 15-28.

Miyake, H., I. Hara, S. Kamidono and M. E. Gleave (2001). "Synergistic chemsensitization and inhibition of tumor growth and metastasis by the antisense oligodeoxynucleotide targeting clusterin gene in a human bladder cancer model." Clinical Cancer Research **7**(12): 4245-4252.

Murakami, Y., S. Kubo, A. Tamori, S. Itami, E. Kawamura, K. Iwaisako, K. Ikeda, N. Kawada, T. Ochiya and Y. Taguchi (2015). "Comprehensive analysis of transcriptome and metabolome analysis in Intrahepatic Cholangiocarcinoma and Hepatocellular Carcinoma." Scientific reports **5**: 16294.

Nagini, S. (2012). "Carcinoma of the stomach: A review of epidemiology, pathogenesis, molecular genetics and chemoprevention." World J Gastrointest Oncol **4**(7): 156-169.

Nesvizhskii, A. I., A. Keller, E. Kolker and R. Aebersold (2003). "A statistical

model for identifying proteins by tandem mass spectrometry." Analytical chemistry **75**(17): 4646-4658.

Otsuji, E., Y. Kuriu, D. Ichikawa, K. Okamoto, T. Ochiai, A. Hagiwara and H. Yamagishi (2004). "Time to death and pattern of death in recurrence following curative resection of gastric carcinoma: analysis based on depth of invasion." World journal of surgery **28**(9): 866-869.

Park, J.-M., J.-H. Park, D.-G. Mun, J. Bae, J. H. Jung, S. Back, H. Lee, H. Kim, H.-J. Jung and H. K. Kim (2015). "Integrated analysis of global proteome, phosphoproteome, and glycoproteome enables complementary interpretation of disease-related protein networks." Scientific reports **5**.

Park, S. H., T. S. Sohn, J. Lee, D. H. Lim, M. E. Hong, K.-M. Kim, I. Sohn, S. H. Jung, M. G. Choi and J. H. Lee (2015). "Phase III trial to compare adjuvant chemotherapy with capecitabine and cisplatin versus concurrent chemoradiotherapy in gastric cancer: final report of the adjuvant chemoradiotherapy in stomach tumors trial, including survival and subset analyses." Journal of Clinical Oncology **33**(28): 3130-3136.

Parker, C. E. and C. H. Borchers (2014). "Mass spectrometry based biomarker discovery, verification, and validation–Quality assurance and control of protein biomarker assays." Molecular oncology **8**(4): 840-858.

Parsonnet, J., G. D. Friedman, D. P. Vandersteen, Y. Chang, J. H. Vogelman, N. Orentreich and R. K. Sibley (1991). "Helicobacter pylori infection and the

risk of gastric carcinoma." New England Journal of Medicine **325**(16): 1127-1131.

Pascal, L. E., R. Z. Vêncio, L. S. Page, E. S. Liebeskind, C. P. Shadle, P. Troisch, B. Marzolf, L. D. True, L. E. Hood and A. Y. Liu (2009). "Gene expression relationship between prostate cancer cells of Gleason 3, 4 and normal epithelial cells as revealed by cell type-specific transcriptomes." BMC cancer **9**(1): 452.

Pavelka, N., M. L. Fournier, S. K. Swanson, M. Pelizzola, P. Ricciardi-Castagnoli, L. Florens and M. P. Washburn (2008). "Statistical similarities between transcriptomics and quantitative shotgun proteomics data." Molecular & Cellular Proteomics **7**(4): 631-644.

Piñeiro, M., M. a. A. Alava, N. González-Ramón, J. Osada, P. Lasierra, L. Larrad, A. Piñeiro and F. n. Lampreave (1999). "ITIH4 serum concentration increases during acute-phase processes in human patients and is up-regulated by interleukin-6 in hepatocarcinoma HepG2 cells." Biochemical and biophysical research communications **263**(1): 224-229.

Pins, M., J. Fiadjoe, F. Korley, M. Wong, A. Rademaker, B. Jovanovic, T. Yoo, J. Kozlowski, A. Raji and X. Yang (2004). "Clusterin as a possible predictor for biochemical recurrence of prostate cancer following radical prostatectomy with intermediate Gleason scores: a preliminary report." Prostate cancer and prostatic diseases **7**(3): 243-248.

Qin, L.-X., Q. Zhou, F. Bogomolniy, L. Villafania, N. Olvera, M. Cavatore, J. M. Satagopan, C. B. Begg and D. A. Levine (2014). "Blocking and randomization to improve molecular biomarker discovery." Clinical Cancer Research **20**(13): 3371-3378.

Qiu, F.-M., J.-K. Yu, Y.-D. Chen, Q.-F. Jin, M.-H. Sui and J. Huang (2009). "Mining novel biomarkers for prognosis of gastric cancer with serum proteomics." Journal of Experimental & Clinical Cancer Research **28**(1): 1.

Rugge, M., G. Busatto, M. Cassaro, Y. H. Shiao, V. Russo, G. Leandro, C. Avellini, A. Fabiano, A. Sidoni and A. Covacci (1999). "Patients younger than 40 years with gastric carcinoma." Cancer **85**(12): 2506-2511.

Schiess, R., B. Wollscheid and R. Aebersold (2009). "Targeted proteomic strategy for clinical biomarker discovery." Molecular oncology **3**(1): 33-44.

Spolverato, G., A. Ejaz, Y. Kim, M. H. Squires, G. A. Poultsides, R. C. Fields, C. Schmidt, S. M. Weber, K. Votanopoulos and S. K. Maithel (2014). "Rates and patterns of recurrence after curative intent resection for gastric cancer: a United States multi-institutional analysis." Journal of the American College of Surgeons **219**(4): 664-675.

Subbannayya, Y., S. A. Mir, S. Renuse, S. S. Manda, S. M. Pinto, V. N. Puttamallesh, H. S. Solanki, H. Manju, N. Syed and R. Sharma (2015). "Identification of differentially expressed serum proteins in gastric adenocarcinoma." Journal of proteomics **127**: 80-88.

Surinova, S., R. Hüttenhain, C.-Y. Chang, L. Espona, O. Vitek and R. Aebersold (2013). "Automated selected reaction monitoring data analysis workflow for large-scale targeted proteomic studies." Nature protocols **8**(8): 1602-1619.

Szklarczyk, D., A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos and K. P. Tsafou (2014). "STRING v10: protein–protein interaction networks, integrated over the tree of life." Nucleic acids research: gku1003.

Takano, S., H. Yoshitomi, A. Togawa, K. Sogawa, T. Shida, F. Kimura, H. Shimizu, T. Tomonaga, F. Nomura and M. Miyazaki (2008). "Apolipoprotein C-1 maintains cell survival by preventing from apoptosis in pancreatic cancer cells." Oncogene **27**(20): 2810-2822.

Tsai, M.-M., C.-S. Wang, C.-Y. Tsai, H.-C. Chi, Y.-H. Tseng and K.-H. Lin (2014). "Potential prognostic, diagnostic and therapeutic markers for human gastric cancer." World J Gastroenterol **20**(38): 13791-13803.

Uen, Y.-H., K.-Y. Lin, D.-P. Sun, C.-C. Liao, M.-S. Hsieh, Y.-K. Huang, Y.-W. Chen, P.-H. Huang, W.-J. Chen and C.-C. Tai (2013). "Comparative proteomics, network analysis and post-translational modification identification reveal differential profiles of plasma Con A-bound glycoprotein biomarkers in gastric cancer." Journal of proteomics **83**: 197-213.

Vanhollebeke, B. and E. Pays (2006). "The function of apolipoproteins L."

Cellular and molecular life sciences **63**(17): 1937-1944.

Vargas, T., J. Moreno-Rubio, J. Herranz, P. Cejas, S. Molina, M. González-Vallinas, R. Ramos, E. Burgos, C. Aguayo and A. B. Custodio (2014). "Genes associated with metabolic syndrome predict disease-free survival in stage II colorectal cancer patients. A novel link between metabolic dysregulation and colorectal cancer." Molecular oncology **8**(8): 1469-1481.

Wang, C., J. Zhang, M. Cai, Z. Zhu, W. Gu, Y. Yu and X. Zhang (2015). "DBGC: A Database of Human Gastric Cancer." PloS one **10**(11): e0142591.

Wang, X., S. Abraham, J. A. McKenzie, N. Jeffs, M. Swire, V. B. Tripathi, U. F. Luhmann, C. A. Lange, Z. Zhai and H. M. Arthur (2013). "LRG1 promotes angiogenesis by modulating endothelial TGF-[bgr] signalling." Nature **499**(7458): 306-311.

Wu, J.-Y., C.-C. Cheng, J.-Y. Wang, D.-C. Wu, J.-S. Hsieh, S.-C. Lee and W.-M. Wang (2014). "Discovery of tumor markers for gastric cancer by proteomics." PloS one **9**(1): e84158.

Wu, J., X. Liu, H. Cai and Y. Wang (2014). "Prediction of tumor recurrence after curative resection in gastric carcinoma based on bcl-2 expression." World journal of surgical oncology **12**(1): 1.

Wu, R.-C., C. Yu, K. Liu, H. Huang, S. Ip, J. Lin, F. Chueh, J. Yang and J. Chung (2012). "Citosol (thiamylal sodium) triggers apoptosis and affects gene expressions of murine leukemia RAW 264.7 cells." Human & experimental

toxicology **31**(8): 771-779.

Xue, A., C. Scarlett, L. Chung, G. Butturini, A. Scarpa, R. Gandy, S. Wilson, R. Baxter and R. C. Smith (2010). "Discovery of serum biomarkers for pancreatic adenocarcinoma using proteomic analysis." British journal of cancer **103**(3): 391-400.

Yamazaki, H., A. Oshima, R. Murakami, S. Endoh and T. Ubukata (1989). "A long-term follow-up study of patients with gastric cancer detected by mass screening." Cancer **63**(4): 613-617.

Yang, G.-D., X.-M. Yang, H. Lu, Y. Ren, M.-Z. Ma, L.-Y. Zhu, J.-H. Wang, W.-W. Song, W.-M. Zhang and R. Zhang (2014). "SERPINA3 promotes endometrial cancer cells growth by regulating G2/M cell cycle checkpoint and apoptosis." International journal of clinical and experimental pathology **7**(4): 1348.

Yom, C. K., H.-Y. Woo, S. Y. Min, S. Y. Kang and H. S. Kim (2009). "Clusterin overexpression and relapse-free survival in breast cancer." Anticancer research **29**(10): 3909-3912.

Yong, C.-S., C.-M. O. Yang, Y.-H. Chou, C.-S. Liao, C.-W. Lee and C.-C. Lee (2012). "CD44/CD24 expression in recurrent gastric cancer: a retrospective analysis." BMC gastroenterology **12**(1): 1.

Yoo, C., S. Noh, D. Shin, S. Choi and J. Min (2000). "Recurrence following curative resection for gastric carcinoma." British Journal of Surgery **87**(2):

236-242.

Younesi, E., L. Toldo, B. Müller, C. M. Friedrich, N. Novac, A. Scheer, M. Hofmann-Apitius and J. Fluck (2012). "Mining biomarker information in biomedical literature." BMC medical informatics and decision making **12**(1): 1.

Zhang, H., J. K. Kim, C. A. Edwards, Z. Xu, R. Taichman and C.-Y. Wang (2005). "Clusterin inhibits apoptosis by interacting with activated Bax." Nature cell biology **7**(9): 909-915.

Zhou, X. M., B. C. Y. Wong, X. M. Fan, H. B. Zhang, M. C. M. Lin, H. F. Kung, D. M. Fan and S. K. Lam (2001). "Non-steroidal anti-inflammatory drugs induce apoptosis in gastric cancer cells through up-regulation of bax and bak." Carcinogenesis **22**(9): 1393-1397.

6. Abstract in Korean

위암은 두 번째로 높은 사망률을 보이는 가장 일반적인 암 중의 하나이다. 최근 임상 치료 요법 발전에도 불구하고, 전체 위암 환자의 재발률은 여전히 높게 남아 있다 (~55%). 따라서 위암 조기 진단 및 예후에 대한 효과적인 임상 치료법을 개발하기 위해서는 위암 재발 기전의 이해가 중요하다.

본 연구는 구축된 프로테오믹스 정량분석법 기술인 다중반응검지법 플랫폼을 이용하여 위암 재발 혈액 바이오마커를 발굴하는 것에 목적을 두었다. 혈액 바이오마커 발굴 플랫폼을 구축하기 위하여, 통합 프로테오믹스 분석법 [(1) 이미 보고된 위암 특이적 혈액 바이오마커 데이터베이스 마이닝, (2) 조기위암 환자 조직시료를 이용한 프로테오믹스 정성 분석, (3) 위암 환자 혈액 시료를 이용한 프로테오믹스 정성 분석]을 이용하여 330개 혈액 단백질을 위암 재발 혈액 바이오마커 1차 후보군으로 선정하였다.

선정된 1차 바이오마커 후보군을 대상으로 선행 다중반응검지법 분석 및 AuDIT 분석을 통해 신뢰성 있게 정량 분석이 가능한 94개의 단백질을 최종 위암 재발 혈액 바이오마커 후보군으로

선정하였다.

다중 바이오마커 패널 구축을 위해, D2 림프절 절제술 후 항암치료를 받은 180명의 위암 환자들을 training set으로 선정하였고, 이를 두 군 (치료 반응 환자군=133명, 재발 환자군=47명)으로 분류하여 다중반응검지법 분석을 수행하였다. 개별 환자들의 다중반응검지법 정량 결과를 추출 후, 통계 분석을 수행하여 6개 단백질 (APOA2, APOC1, CLU, ITIH4, LRG1, SERPINA3)을 다중 바이오마커 패널로 구축하였다. 이 패널은 두 군에서 통계적 유의미한 값 (p-value < 0.05)을 가지며 유의미한 발현차이를 보였으며, 0.810의 AUC 값 및 높은 예측도를 나타내었다 (치료 반응 환자군=95.5%, 재발 환자군=61.7%).

구축된 다중 바이오마커 패널을 검증하기 위해 독립적인 64명의 위암 환자 (치료 반응 환자군=43명, 재발 환자군=21명)를 test set으로 선정하여 다중반응검지법 분석 및 통계 분석을 수행하였다. Training set 결과와 비교하였을 때, 6개 다중 바이오마커 혈액 단백질이 test set 내에서도 통계적 유의미한 값 (p-value < 0.05)을 가지며 두 군에서 동일한 발현 경향성을 보이는 것을 확인하였다.

결론적으로, 본 연구를 통해 선정된 6개 단백질들이 위암 환자의 재발을 확인하는 진단 시그니처로 사용될 수 있으며, 다중반응검지법 기반 혈액 바이오마커 구축 플랫폼이 임상적 바이오마커 발굴 및 검증 프로세스에서 유용한 방법이 될 수 있을 것이라고 사료되었다.

주요어: 위암, 재발, 다중반응검지법, 다중 바이오마커 패널

학번: 2013-30836